Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder

ADHD Working Grp Psychiat Genomics

2019-01

ADHD Working Grp Psychiat Genomics , Early Lifecourse Genetic , 23andMe Res Team & Daly , M J 2019 , ' Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder ' , Nature Genetics , vol. 51 , no. 1 , pp. 63-+ . https://doi.org/10.1038/s41588-018-0269-

http://hdl.handle.net/10138/309285
https://doi.org/10.1038/s41588-018-0269-7

unspecified
publishedVersion

nature genetics

ARTICLES

https://doi.org/10.1038/s41588-018-0269-7

Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder

Ditte Demontis[1,2,3,69], Raymond K. Walters[4,5,69], Joanna Martin[5,6,7], Manuel Mattheisen[1,2,3,8,9,10], Thomas D. Als[1,2,3], Esben Agerbo[1,11,12], Gísli Baldursson[13], Rich Belliveau[5], Jonas Bybjerg-Grauholm[1,14], Marie Bækvad-Hansen[1,14], Felecia Cerrato[5], Kimberly Chambert[5], Claire Churchhouse[4,5,15], Ashley Dumont[5], Nicholas Eriksson[16], Michael Gandal[17,18,19,20], Jacqueline I. Goldstein[4,5,15], Katrina L. Grasby[21], Jakob Grove[1,2,3,22], Olafur O. Gudmundsson[13,23,24], Christine S. Hansen[1,14,25], Mads Engel Hauberg[1,2,3], Mads V. Hollegaard[1,14], Daniel P. Howrigan[4,5], Hailiang Huang[4,5], Julian B. Maller[5,26], Alicia R. Martin[4,5,15], Nicholas G. Martin[21], Jennifer Moran[5], Jonatan Pallesen[1,2,3], Duncan S. Palmer[4,5], Carsten Bøcker Pedersen[1,11,12], Marianne Giørtz Pedersen[1,11,12], Timothy Poterba[4,5,15], Jesper Buchhave Poulsen[1,14], Stephan Ripke[4,5,27], Elise B. Robinson[4,28], F. Kyle Satterstrom[4,5,15], Hreinn Stefansson[23], Christine Stevens[5], Patrick Turley[4,5], G. Bragi Walters[23,24], Hyejung Won[17,18], Margaret J. Wright[29], ADHD Working Group of the Psychiatric Genomics Consortium (PGC)[30], Early Lifecourse & Genetic Epidemiology (EAGLE) Consortium[30], 23andMe Research Team[30], Ole A. Andreassen[31], Philip Asherson[32], Christie L. Burton[33], Dorret I. Boomsma[34,35], Bru Cormand[36,37,38,39], Søren Dalsgaard[11], Barbara Franke[40], Joel Gelernter[41,42], Daniel Geschwind[17,18,19], Hakon Hakonarson[43], Jan Haavik[44,45], Henry R. Kranzler[46,47], Jonna Kuntsi[32], Kate Langley[7,48], Klaus-Peter Lesch[49,50,51], Christel Middeldorp[34,52,53], Andreas Reif[54], Luis Augusto Rohde[55,56], Panos Roussos[57,58,59,60], Russell Schachar[33], Pamela Sklar[57,58,59], Edmund J. S. Sonuga-Barke[61], Patrick F. Sullivan[6,62], Anita Thapar[7], Joyce Y. Tung[16], Irwin D. Waldman[63], Sarah E. Medland[21], Kari Stefansson[23,24], Merete Nordentoft[1,64], David M. Hougaard[1,14], Thomas Werge[1,25,65], Ole Mors[1,66], Preben Bo Mortensen[1,2,11,12], Mark J. Daly[4,5,15,67], Stephen V. Faraone[68,70*], Anders D. Børglum[1,2,3,70*] and Benjamin M. Neale[4,5,15,70*]

Attention deficit/hyperactivity disorder (ADHD) is a highly heritable childhood behavioral disorder affecting 5% of children and 2.5% of adults. Common genetic variants contribute substantially to ADHD susceptibility, but no variants have been robustly associated with ADHD. We report a genome-wide association meta-analysis of 20,183 individuals diagnosed with ADHD and 35,191 controls that identifies variants surpassing genome-wide significance in 12 independent loci, finding important new information about the underlying biology of ADHD. Associations are enriched in evolutionarily constrained genomic regions and loss-of-function intolerant genes and around brain-expressed regulatory marks. Analyses of three replication studies: a cohort of individuals diagnosed with ADHD, a self-reported ADHD sample and a meta-analysis of quantitative measures of ADHD symptoms in the population, support these findings while highlighting study-specific differences on genetic overlap with educational attainment. Strong concordance with GWAS of quantitative population measures of ADHD symptoms supports that clinical diagnosis of ADHD is an extreme expression of continuous heritable traits.

ADHD is a neurodevelopmental psychiatric disorder that affects around 5% of children and adolescents and 2.5% of adults worldwide[1]. ADHD is often persistent and markedly impairing, with increased risk of harmful outcomes, such as injuries[2], traffic accidents[3], increased healthcare utilization[4,5], substance abuse[6], criminality[7], unemployment[8], divorce[4], suicide[9], AIDS risk behaviors[8] and premature mortality[10]. Epidemiologic and clinical studies implicate genetic and environmental risk factors that affect the structure and functional capacity of brain networks involved in behavior and cognition[1] in the etiology of ADHD.

A full list of affiliations appears at the end of the paper.

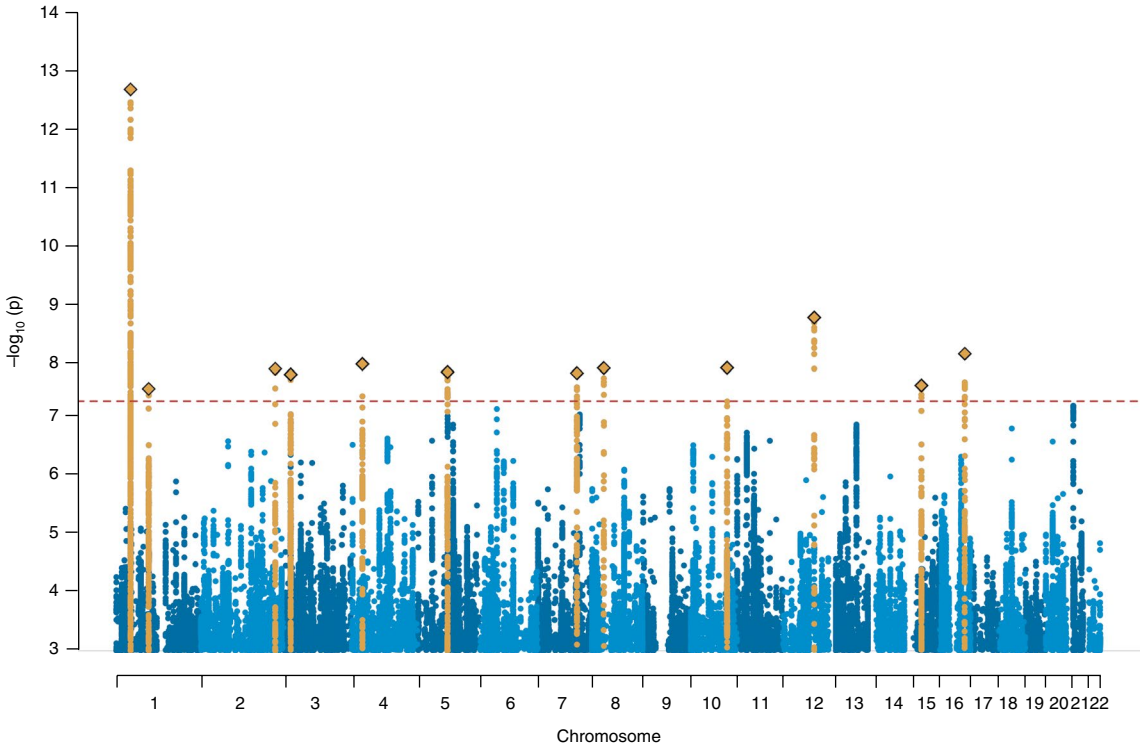

Fig. 1 | Manhattan plot of the results from the GWAS meta-analysis of ADHD. The index variants in the 12 genome-wide significant loci are highlighted as an orange diamond. Index variants located with a distance <400 kb are considered as one locus. The *y* axis represents −log(two-sided *P* values) for association of variants with ADHD, from meta-analysis using an inverse-variance weighted fixed effects model and a total sample size of 20,183 individuals with ADHD and 35,191 controls. The horizontal red line represents the threshold for genome-wide significance.

Consensus estimates from more than 30 twin studies indicate that the heritability of ADHD is 70–80% throughout the lifespan[11,12] and that environmental risks are those not shared by siblings[13]. Twin studies also suggest that diagnosed ADHD represents the extreme tail of one or more heritable quantitative traits[14]. Additionally, family and twin studies report genetic overlap between ADHD and other conditions, including antisocial personality disorder/behaviors[15], cognitive impairment[16], autism spectrum disorder[17,18], schizophrenia[19], bipolar disorder[20], and major depressive disorder[21].

Thus far, genome-wide association studies (GWASs) to identify common DNA variants that increase the risk of ADHD have not been successful[22]. Nevertheless, genome-wide SNP heritability estimates range from 0.10–0.28 (ref. [23,24]), supporting the notion that common variants comprise a significant fraction of the risk underlying ADHD[25] and that with increasing sample size, and thus, increasing statistical power, genome-wide significant loci will emerge.

Previous studies have demonstrated that the common variant risk, also referred to as the SNP heritability, of ADHD is also associated with depression[25], conduct problems[26], schizophrenia[27], continuous measures of ADHD symptoms[28,29] and other neurodevelopmental traits[29] in the population. Genetic studies of quantitative ADHD symptom scores in children further support the hypothesis that ADHD is the extreme of a quantitative trait[30].

Here, we present a genome-wide meta-analysis identifying the first genome-wide significant loci for ADHD using a combined sample of 55,374 individuals from an international collaboration. We also strengthen the case that the clinical diagnosis of ADHD is the extreme expression of one or more heritable quantitative traits, at least as it pertains to common variant genetic risk, by integrating our results with previous GWASs of ADHD-related behavior in the general population.

Results

Genome-wide significantly associated ADHD risk loci. Genotype array data for 20,183 individuals with ADHD and 35,191 controls were collected from 12 cohorts (Supplementary Table 1). These samples included a population-based cohort of 14,584 individuals with ADHD and 22,492 controls from Denmark collected by the Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH; Supplementary Fig. 1), and 11 European, North American and Chinese cohorts aggregated by the Psychiatric Genomics Consortium (PGC). Individuals with ADHD in iPSYCH were identified from the national Psychiatric Central Research Register and diagnosed by psychiatrists at a psychiatric hospital according to ICD10 (F90.0) and then genotyped using Illumina PsychChip. Designs for the PGC cohorts have been described previously[22,24,25,31,32] (detailed cohort descriptions in Supplementary Note). All relevant ethical permissions and informed consent were obtained for the included cohorts (details in approval authorities in Supplementary Note).

Prior to analysis, stringent quality control procedures were performed on the genotyped markers and individuals in each cohort using a standardized pipeline[33] (Methods). Related individuals were removed, and genetic outliers within each cohort were excluded based on principal component analysis. Non-genotyped markers were imputed using the 1000 Genomes Project Phase 3 reference panel[34] (Methods).

GWAS was conducted in each cohort using logistic regression with the imputed additive genotype dosages. Principal components were included as covariates to correct for population stratification[35] (Supplementary Note), and variants with imputation INFO score < 0.8 or minor allele frequency (MAF) < 0.01 were excluded. The GWASs were then meta-analyzed using an inverse-variance weighted fixed effects model[36]. The single Chinese cohort included

Table 1 | Results for the genome-wide significant index variants in the 12 loci associated with ADHD identified in the GWAS meta-analysis of 20,183 individuals with ADHD and 35,191 controls

Locus	Chr	BP	Index variant	Genes	A1	A2	A1 freq	OR	*P* value
1	1	44184192	rs11420276	*ST3GAL3, KDM4A, KDM4A-AS1, PTPRF, SLC6A9, ARTN, DPH2, ATP6V0B, B4GALT2, CCDC24, IPO13*	G	GT	0.696	1.113	2.14×10^{-13}
2	1	96602440	rs1222063	Intergenic	A	G	0.328	1.101	3.07×10^{-8}
3	2	215181889	rs9677504	*SPAG16*	A	G	0.109	1.124	1.39×10^{-8}
4	3	20669071	rs4858241	Intergenic	T	G	0.622	1.082	1.74×10^{-8}
5	4	31151456	rs28411770	*PCDH7, LINC02497*	T	C	0.651	1.090	1.15×10^{-8}
6	5	87854395	rs4916723	*LINC00461, MIR9-2, LINC02060, TMEM161B-AS1*	A	C	0.573	0.926	1.58×10^{-8}
7	7	114086133	rs5886709	*FOXP2, MIR3666*	G	GTC	0.463	1.079	1.66×10^{-8}
8	8	34352610	rs74760947	*LINC01288*	A	G	0.957	0.835	1.35×10^{-8}
9	10	106747354	rs11591402	*SORCS3*	A	T	0.224	0.911	1.34×10^{-8}
10	12	89760744	rs1427829	*DUSP6, POC1B*	A	G	0.434	1.083	1.82×10^{-9}
11	15	47754018	rs281324	*SEMA6D*	T	C	0.531	0.928	2.68×10^{-8}
12	16	72578131	rs212178	*LINC01572*	A	G	0.883	0.891	7.68×10^{-9}

Index variants are LD independent ($r^2 < 0.1$) and are merged into one locus when located with a distance <400 kb. The location (chromosome (chr) and base position (BP)), alleles (A1 and A2), allele frequency (A1 freq), odds ratio (OR) of the effect with respect to A1 and association *P* values from inverse-variance weighted fixed effects model of the index variant are given, along with genes within 50 kb of the credible set for the locus.

had insufficient sample size for well-powered transethnic modeling (Supplementary Fig. 2). Association results were considered only for variants with an effective sample size >70% of the full meta-analysis, leaving 8,047,421 variants in the final meta-analysis. A meta-analysis restricted to individuals of European ancestry (19,099 with ADHD, 34,194 controls) was also performed to facilitate secondary analyses (Supplementary Note).

In total, 304 genetic variants in 12 loci surpassed the threshold for genome-wide significance ($P < 5 \times 10^{-8}$; Fig. 1, Table 1 and Supplementary Fig. 3). Results for the European ancestry meta-analysis were substantively similar (Supplementary Fig. 4). No marker demonstrated significant heterogeneity between studies (Supplementary Figs. 5 and 6), and no heterogeneity was observed between the Chinese and European ancestry cohorts (Supplementary Fig. 2). Conditional analysis within each locus did not identify any independent secondary signals meeting genome-wide significance (Methods; Supplementary Table 2).

Homogeneity of effects between cohorts. No genome-wide significant heterogeneity was observed in the ADHD GWAS meta-analysis (Supplementary Note). A genetic correlation analysis (Methods) provided further evidence that effects were consistent across cohort study designs. The estimated genetic correlation between the European ancestry PGC samples and the iPSYCH sample from linkage disequilibrium (LD) score regression[37] was not significantly less than 1 ($r_g = 1.17$, standard error (SE) = 0.20). The correlation between European ancestry PGC case/control and trio cohorts estimated with bivariate GREML was similarly close to 1 ($r_g = 1.02$, SE = 0.32; Supplementary Table 3).

Polygenic risk scores (PRSs)[38] were also consistent across target samples. PRSs computed in each PGC study using iPSYCH as the training sample were consistently higher in ADHD compared with controls or pseudocontrols (Supplementary Fig. 7). Increasing deciles of PRS in the PGC were associated with a higher odds ratio (OR) for ADHD (Fig. 2). A similar pattern was seen in five-fold cross-validation in the iPSYCH cohort, with PRS for each subset computed from the other four iPSYCH subsets and the PGC samples used as training samples (Methods; Fig. 2). Across iPSYCH subsets, the mean of the maximum variance explained by the estimated PRS (Nagelkerke's R^2) was 5.5% (SE = 0.0012) (Supplementary Fig. 8). The difference in standardized PRS between cases and controls was stable across iPSYCH subsets (OR = 1.56, 95% confidence interval (CI): 1.53–1.60; Supplementary Fig. 9) and across waves and PGC cohorts (Supplementary Fig. 10). These results further support the notion of highly polygenic architecture of ADHD and demonstrate that ADHD risk is significantly associated with PRS in a dose-dependent manner.

Polygenic architecture of ADHD. To assess the proportion of phenotypic variance explained by common variants, we applied LD score regression[37] to results from the European ancestry meta-analysis (Methods). Assuming a population prevalence of 5% for ADHD[39], we estimated the liability-scale SNP heritability as $h^2_{SNP} = 0.216$ (SE = 0.014, $P = 8.18 \times 10^{-54}$; Supplementary Table 4). These estimated polygenic effects account for 88% (SE = 0.0335) of observed genome-wide inflation of the test statistics in the meta-analysis ($\lambda = 1.200$; quantile-quantile plots in Supplementary Fig. 11); the remaining inflation, which may reflect confounding factors, such as cryptic relatedness and population stratification, is significant but modest (intercept = 1.0362, SE = 0.0099, $P = 2.27 \times 10^{-4}$).

To further characterize the patterns of heritability from the genome-wide association data, we partitioned SNP heritability by functional annotations, as described in Finucane et al.[40], using partitioned LD score regression (Methods). The analysis found significant enrichment in the heritability from SNPs located in conserved regions ($P = 8.49 \times 10^{-10}$; Supplementary Fig. 12), supporting their biological importance. Enrichment of the SNP heritability in cell-type-specific regulatory elements was evaluated using the cell-type-specific group annotations described in Finucane et al.[40]. We observed a significant enrichment of the average per SNP heritability for variants located in central nervous system–specific regulatory elements (enrichment = 2.44, SE = 0.35, $P = 5.81 \times 10^{-5}$; Supplementary Figs. 13 and 14).

Genetic correlation with other traits. Pairwise genetic correlation with ADHD was estimated for 219 phenotypes using LD score regression[41,42] (Methods; Supplementary Data 1). Forty-three phenotypes demonstrated significant genetic overlap with ADHD

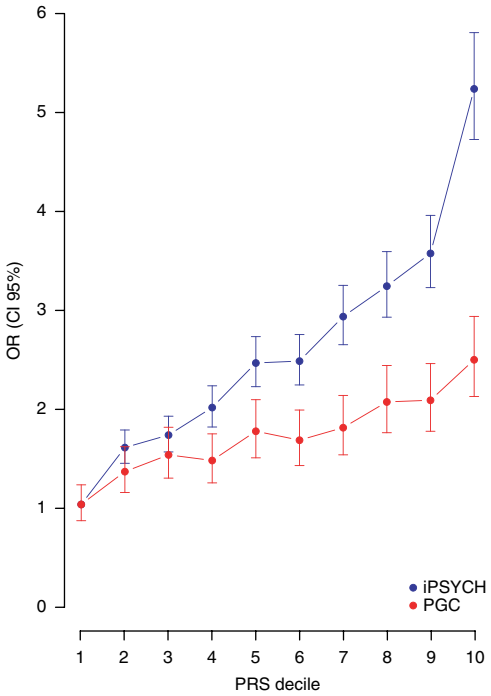

Fig. 2 | Odds ratio by PRS for ADHD. OR by PRS within each decile estimated for $n=18,298$ biologically independent individuals in the PGC samples (red dots) and in $n=37,076$ biologically independent individuals in the iPSYCH sample (blue dots). PRS in the iPSYCH sample were obtained by five leave-one-out analyses, using 4 of 5 groups as training datasets for estimation of SNP weights while estimating PRS for the remaining target group. ORs and 95% confidence limits (error bars) were estimated using logistic regression on the continuous scores.

($P<2.28\times10^{-4}$), including major depressive disorder[43], anorexia nervosa[44], educational outcomes[45–49], obesity-related phenotypes[50–55], smoking[56–58], reproductive success[59], insomnia[60], and mortality[61] (Fig. 3 and Supplementary Table 5). In most domains, the genetic correlation is supported by GWAS of multiple related phenotypes. For the positive genetic correlation with major depressive disorder ($r_g=0.42$, $P=7.38\times10^{-38}$), we also observed a positive correlation with depressive symptoms ($r_g=0.45$, $P=7.00\times10^{-19}$), neuroticism ($r_g=0.26$, $P=1.02\times10^{-8}$) and a negative correlation with subjective well-being ($r_g=-0.28$, $P=3.73\times10^{-9}$). The positive genetic correlations with ever having smoked ($r_g=0.48$, $P=4.33\times10^{-16}$) and with number of cigarettes smoked per day ($r_g=0.45$, $P=1.07\times10^{-5}$) are reinforced by significant positive correlation with lung cancer ($r_g=0.39$, $P=6.35\times10^{-10}$). Similarly, genetic correlations related to obesity include significant relationships with body mass index (BMI; $r_g=0.26$, $P=1.68\times10^{-15}$), waist-to-hip ratio ($r_g=0.30$, $P=1.16\times10^{-17}$), childhood obesity ($r_g=0.22$, $P=3.29\times10^{-6}$), HDL cholesterol ($r_g=-0.22$, $P=2.44\times10^{-7}$), and type 2 diabetes ($r_g=0.18$, $P=7.80\times10^{-5}$). Additionally the negative correlation with years of schooling ($r_g=-0.53$, $P=6.02\times10^{-80}$) is supported by a negative genetic correlation with human intelligence ($r_g=-0.41$, $P=7.03\times10^{-26}$). Finally, the genetic correlation with reproduction includes a negative correlation with age of first birth ($r_g=-0.612$, $P=3.70\times10^{-61}$) and a positive correlation with number of children ever born ($r_g=0.42$, $P=8.51\times10^{-17}$).

Biological annotation of significant loci. For the 12 genome-wide significant loci, Bayesian credible sets were defined to identify the set of variants at each locus most likely to include a variant with causal effect (Methods, Supplementary Data 2 and Supplementary Table 6). Biological annotations of the variants in the credible set were then considered to identify functional or regulatory variants, common chromatin marks, and variants associated with gene expression (eQTLs) or in regions with gene interactions observed in Hi-C data (Methods; Supplementary Data 3). Broadly, the significant loci do not coincide with candidate genes proposed to play a role in ADHD[62].

Here, we highlight genes that are identified in the regions of association (also Supplementary Table 7). The loci on chromosomes 2, 7 and 10 each have credible sets localized to a single gene with limited additional annotations. In the chromosome 7 locus, *FOXP2* encodes a forkhead/winged-helix transcription factor and is known to play an important role in synapse formation and neural mechanisms mediating the development of speech and learning[63–65]. Comorbidity of ADHD with specific developmental disorders of language and learning is common (7–11%)[66,67], and poor language skills have been associated with higher inattention or hyperactivity symptoms in primary school[68]. On chromosome 10, the ADHD association is intronic, located in *SORCS3*, which encodes a brain-expressed transmembrane receptor that is important for neuronal development and plasticity[69] and has previously been associated with depression[43,70].

Genome-wide significant loci on chromosomes 12 and 15 have more biological annotations supporting the colocalized genes. The credible set on chromosome 12 spans *DUSP6* and includes an annotated missense variant in the first exon and an insertion near the transcription start site, though neither is the lead variant in the locus (Supplementary Data 4). *DUSP6* encodes a dual specificity phosphatase[71] and may play a role in regulating neurotransmitter homeostasis by affecting dopamine levels in the synapses[72,73]. Regulation of dopamine levels is likely to be relevant to ADHD, as widely used ADHD medications have dopaminergic targets[74,75] that increase the availability of synaptic dopamine. The chromosome 15 locus is located in *SEMA6D*, and the majority of variants in the credible set are strongly associated with expression of *SEMA6D* in fibroblasts[76]. *SEMA6D* is active in the brain during embryonic development and may play a role in neuronal wiring[77]. Furthermore, variants in *SEMA6D* have previously been associated with educational attainment[78].

Credible set annotations at the remaining loci are more diverse (Supplementary Data 3). The most strongly associated locus on chromosome 1 (index variant rs112984125) covers a gene-rich 250-kb region of strong LD. The index variant is intronic to *ST3GAL3*, and most SNPs in the credible set are strongly associated with expression of *ST3GAL3* in whole blood[79] (Supplementary Data 3). Missense mutations in *ST3GAL3* have been shown to cause autosomal recessive intellectual disability[80]. Hi-C and eQTL annotations suggest multiple alternative genes, however, including *PTPRF* (Supplementary Data 4). The locus also includes an intergenic variant, rs11210892, that has previously been associated with schizophrenia[33].

On chromosome 5, the credible set includes links to *LINC00461* and *TMEM161B* (Supplementary Data 3). The function of *LINC00461* is unclear, but the RNA has highly localized expression in the brain[81], and the genome-wide significant locus overlaps with variants in *LINC00461* associated with educational attainment[78]. Alternatively, a genome-wide significant SNP in this locus (rs304132) is located in *MEF2C-AS1*, of strong interest given previous associations between *MEF2C* and severe intellectual disability[82–84], cerebral malformation[83], depression[70], schizophrenia[33] and Alzheimer's disease[85], but the corresponding variant is not supported by the credible set analysis. Credible set annotations for other significant loci are similarly cryptic.

Analysis of gene sets. Competitive gene-based tests were performed for *FOXP2* target genes, highly constrained genes and for all Gene Ontology terms[86] from MsigDB 6.0 (ref. [87]) using MAGMA[88] (Methods). Association results for individual genes are consistent

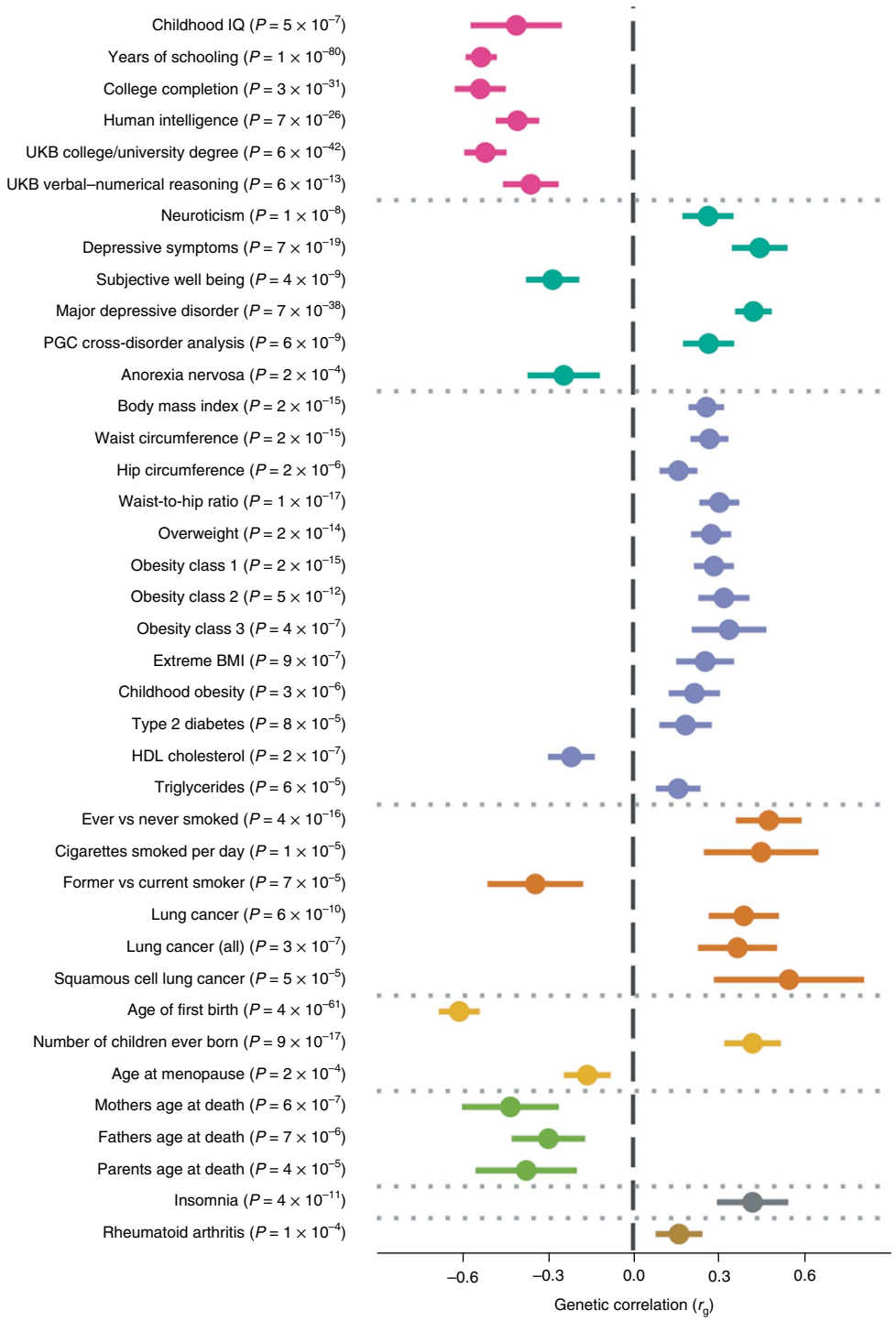

Fig. 3 | Genetic correlations of ADHD with other phenotypes. Significant genetic correlations between ADHD (results from European GWAS meta-analysis of 19,099 individuals with ADHD, 34,194 controls) and other traits reveal overlap of genetic risk factors for ADHD across several groups of traits (grouping indicated by a vertical line): educational, psychiatric or personality, weight (and possible weight-related traits), smoking behavior or smoking-related cancer, reproductive traits and parental longevity (sample size of the external GWASs are presented in Supplementary Table 5). In total, 219 traits were tested, and only traits significant after Bonferroni correction are presented. Results are omitted for significant correlations with two previous GWAS of years of schooling and two GWAS whose the discovery sample was not restricted to European ancestry. Genetic correlation is presented as a dot and error bars indicate 95% confidence limits.

with the genome-wide significant loci for the GWAS (Supplementary Table 8); however, four new genes passed the threshold for exome-wide significant association (Supplementary Fig. 15a–d). Three independent sets of *FOXP2* downstream target genes[89,90] were tested (Methods), none of which demonstrated significant association to ADHD (Supplementary Table 9). The lack of association might be caused by unknown functions of *FOXP2* driving ADHD risk, insufficient power to detect relevant downstream genes or because only a small subset of biological functions regulated by FOXP2 are relevant to ADHD pathogenesis.

Consistent with the partitioning of heritability, a set of 2,932 genes that are highly constrained and show high intolerance to loss of function[91] showed significant association with ADHD ($\beta = 0.062$, $P = 2.6 \times 10^{-4}$; Supplementary Table 10). We also found little evidence for effects in previously proposed candidate genes for ADHD[62]; of the nine proposed genes, only *SLC9A9* showed weak association with ADHD ($P = 3.4 \times 10^{-4}$; Supplementary Table 11). None of the Gene Ontology gene sets were significant after correcting for multiple testing, although the most associated included interesting nominally significant pathways such as 'dopamine receptor binding' ($P = 0.0010$) and 'excitatory synapse' ($P = 0.0088$; Supplementary Data 5).

Replication of GWAS loci. For replication, we evaluated the comparison of the GWAS meta-analysis of ADHD with three other independent ADHD-related GWASs: replication of top loci in an Icelandic cohort with ADHD status derived from medical records of ICD codes and medication history by deCODE (5,085 with ADHD, 131,122 controls), a GWAS of self-reported ADHD status among 23andMe research participants (5,857 with ADHD, 70,393 controls) and a meta-analysis of GWAS of childhood rating scales of ADHD symptoms performed by the EAGLE consortium (17,666 children <13 years of age)[30] and QIMR[92] (2,798 adolescents), referred to as EAGLE/QIMR hereafter. Although the phenotyping and cohort ascertainment of the 23andMe and EAGLE/QIMR studies differ from the PGC and iPSYCH ADHD meta-analysis (Supplementary Note), they have clear relevance to understanding how the ADHD GWAS results generalize to closely related phenotypes.

Top loci from the ADHD GWAS showed moderate concordance across the three replication studies. Sign concordance between each of the three replication cohorts and the ADHD GWAS was significantly greater than what would be expected by chance (range 72–82% concordant; $P < 0.0167 = 0.05/3$ replication cohorts; Supplementary Table 12) for nominally associated loci from the ADHD GWAS ($P < 1 \times 10^{-6}$), with the highest concordance observed in EAGLE/QIMR. The deCODE and 23andMe results also permit direct comparisons of the magnitude of effect sizes for the top loci in the ADHD GWAS (Supplementary Table 13). Regressing effect size estimates from each replication cohort on estimates from the ADHD GWAS adjusted for winner's curse yields significantly positive slopes (deCODE slope = 0.664, $P = 1.2 \times 10^{-4}$; 23andMe slope = 0.417, $P = 1.11 \times 10^{-3}$), although these slopes are less than one, suggesting imperfect replication. Among the genome-wide significant loci, rs9677504 (*SPAG16* locus) in deCODE and rs112984125 (*ST3GAL3/PTPRF* locus) and rs212178 (*LINC01572* locus) in 23andMe are notable outliers with weak replication results (Methods; Supplementary Figs. 16 and 17).

The genome-wide data available from 23andMe and EAGLE/QIMR showed similar trends for replication. The genetic correlation between EAGLE/QIMR and the ADHD GWAS was extremely strong ($r_g = 0.970$, SE = 0.207, $P = 2.66 \times 10^{-6}$) and not significantly different from 1 (one-sided $P = 0.442$). Genetic correlation with the 23andMe results was weaker but still strongly positive ($r_g = 0.653$, SE = 0.114, $P = 1.11 \times 10^{-8}$), although also significantly less than 1 (one-sided $P = 1.17 \times 10^{-3}$). To explore this lower correlation, we evaluated the genetic correlation between 23andMe and traits from LD Hub (see URLs)[42] to potentially identify differences in the profile of genetic correlations compared with the ADHD GWAS (Methods). This comparison identified striking differences (Supplementary Table 14), most notably that the 23andMe GWAS shows little to no genetic correlation with college completion ($r_g = 0.056$, compared with $r_g = -0.54$ for the primary ADHD GWAS; approximate $P = 1.1 \times 10^{-9}$ for difference) and other education-related phenotypes. Genetic correlations with obesity-related phenotypes were similarly smaller for the 23andMe cohort. The domains in which 23andMe exhibited a trend toward stronger genetic correlations were schizophrenia ($r_g = 0.27$ vs. $r_g = 0.12$ in ADHD, $P = 0.053$) and bipolar disorder ($r_g = 0.029$ vs. $r_g = 0.095$ in ADHD, $P = 0.09$), although these trends are not significant with the approximated test of the difference in genetic correlation.

Finally, we meta-analyzed the ADHD GWAS with each replication cohort. For EAGLE/QIMR, we developed a novel model to meta-analyze the GWAS of the continuous measure of ADHD with the clinical diagnosis in the ADHD GWAS. In brief, we perform a *z*-score based meta-analysis using a weighting scheme derived from the SNP heritability and effective sample size for each phenotype that fully accounts for the differences in measurement scale (detailed description in Supplementary Note and Supplementary Figs. 18–20). This calibration based on the genome-wide estimate of heritability prevents joint meta-analysis of all replication cohorts because genome-wide data is not available for the deCODE study.

Meta-analyses of the ADHD GWAS with each replication study identified ten genome-wide significant loci ($P < 5 \times 10^{-8}$, without multiple testing correction) in meta-analysis with deCODE, ten significant loci with 23andMe, and 15 significant loci with EAGLE/QIMR (Supplementary Data 6 and Supplementary Figs. 21,22). Of the 12 significant loci from the primary ADHD GWAS, four were significant in all three of these replication meta-analyses: index variants rs11420276 (*ST3GAL3/PTPRF*), rs5886709 (*FOXP2*), rs11591402 (*SORCS3*), and rs1427829 (intergenic). The remaining loci were all significant in at least one of the replication meta-analyses. Additionally, ten novel loci reached genome-wide significance in the replication meta-analyses, of which three loci were significant in two of these analyses (Supplementary Data 6): index variants rs1592757/rs30266 (RefSeq *LOC105379109*), rs28452470/rs1443749 (*CADPS2*), and rs2243638/rs9574218 (*RNF219-AS1*). The *CADPS2* locus has recently been identified in autism spectrum disorder as a novel locus shared with educational attainment[93].

Meta-analysis with the 23andMe cohort also found genome-wide significant heterogeneity at the lead chromosome 1 locus from the ADHD GWAS meta-analysis (rs12410155: $I^2 = 97.2$, $P = 2.29 \times 10^{-9}$; Supplementary Figs. 23 and 24). This heterogeneity is consistent with the moderate sign concordance, effect size replication and genetic correlation of the 23andMe cohort with the ADHD GWAS. Notably, the lead chromosome 1 locus in the ADHD GWAS overlaps a reported association with educational attainment[78], suggesting that this heterogeneity is consistent with the much weaker genetic correlation between the 23andMe results and published GWAS of education-related outcomes. No genome-wide significant heterogeneity was observed in the replication meta-analyses with deCODE or EAGLE/QIMR (Supplementary Figs. 25 and 26 and Supplementary Data 6).

Discussion

Our GWAS meta-analysis of ADHD identified the first genome-wide significant risk loci and indicates an important role for common variants in the polygenic architecture of ADHD. Several of the loci are located in or near genes that implicate neurodevelopmental processes that are likely to be relevant to ADHD, including *FOXP2*, *SORCS3* and *DUSP6*. Future work might focus on refining the source of the strong association in each locus, especially the lead locus on chromosome 1, which is complicated by broad LD and substantial heterogeneity between the main meta-analysis of ADHD and the analysis of self-reported ADHD status in 23andMe.

The 12 significant loci are compelling, but only capture a tiny fraction of common variant risk for ADHD. The ORs for the risk increasing allele at the index SNPs in the 12 significant loci are modest, ranging from 1.077 to 1.198 (Table 1). This is within the range of effect sizes for common genetic variants that has been observed for other highly polygenic psychiatric disorders, for example, schizophrenia[33]. A considerably larger proportion of the heritability

of ADHD can be explained by all common variants (h^2_{SNP} = 0.22, SE = 0.01). This is consistent with previous estimates of h^2_{SNP} for ADHD in smaller studies (h^2_{SNP}: 0.1–0.28)[23,24] and also comparable to SNP heritability estimates for schizophrenia (h^2_{SNP} 0.23–0.26)[23,24]. As would be hypothesized for a psychiatric disorder, these effects are enriched in conserved regions and regions containing enhancers and promoters of expression in central nervous system tissues, consistent with previous observations in schizophrenia and bipolar disorder[40]. On the other hand, we do not observe substantial effects in most previously reported candidate genes for ADHD[62].

Along with polygenicity, selection and evolutionary pressures might be an important feature of the architecture of ADHD genetics. We observe that ADHD risk variants are strongly enriched in genomic regions conserved in mammals[94], and constrained genes likely to be intolerant to loss-of-function mutations[91] are associated with ADHD. We also find that common variant risk for ADHD is genetically correlated with having children younger and having more children, in line with epidemiological findings of increased risky sexual behavior[95–97] and increased risk of ADHD for children born to young parents[98–100]. Given the phenotypic[101,102] and genetic[103] correlation of ADHD with reduced educational attainment, positive selective pressure on the genetics of ADHD would be consistent with recently published work suggesting that variants associated with educational attainment are under negative selection in Iceland[104]. Future studies of fecundity and the role of rare and de novo variants in ADHD might provide more insight on selective pressures in ADHD-associated loci.

The observed genetic correlations with educational outcomes and other phenotypes suggest a strong genetic component to the epidemiological correlates of ADHD. The significant positive genetic correlation of ADHD with major depressive disorder and depressive symptoms supports previous findings that suggest a positive genetic overlap between those phenotypes[24,42], as well as the broader genetic overlap of psychiatric disorders[23,24]. Positive genetic correlations between ADHD and health risk behaviors such as smoking and obesity are consistent with the observed increase in those behaviors among individuals with ADHD[105–108] and are indicative of a shared genetic basis for these traits. We also observed a positive genetic correlation of ADHD with insomnia, consistent with reports of sleep disturbances in ADHD[109], but this relationship does not appear to generalize to other sleep-related phenotypes.

These genetic correlations might not generalize to all settings. We observed much weaker genetic correlation of the 23andMe ADHD results with educational attainment, with only partial genetic correlation between 23andMe and the current ADHD GWAS, including significant heterogeneity in the lead chromosome 1 locus. The pattern of replication for the top loci in the deCODE study is stronger but still mixed. These differences may reflect dissimilarities in phenotyping (for example self-report vs. medical records), exclusion of individuals with comorbid psychiatric disorders (deCODE), study population (for example, higher average education and socioeconomic status among 23andMe research participants possibly underrepresenting the proportion of individuals with ADHD with poor educational outcomes in the general population) or other study factors that should be a focus of future work.

On the other hand, the replication results from EAGLE[30]/QIMR[92] are much stronger and support the hypothesis that ADHD is the extreme expression of one or more heritable quantitative traits[110]. We observe strong concordance between the GWAS of ADHD and the previous GWASs of ADHD-related traits in the population, both in terms of genome-wide genetic correlation and concordance at individual loci. Polygenic risk for ADHD has previously been associated with inattentive and hyperactive/impulsive trait variation below clinical thresholds in the population[29]. Shared genetic risk with health risk behaviors may similarly be hypothesized to reflect an impaired ability to self-regulate and inhibit impulsive behavior[111,112]. The observed negative correlation between ADHD and anorexia nervosa might also be related to these behavioral factors.

In summary, we report 12 independent genome-wide significant loci associated with ADHD in a GWAS meta-analysis of 55,374 individuals from 12 study cohorts. The GWAS meta-analysis implicates *FOXP2* and other biologically informative genes as well as constrained regions of the genome as important contributors to the etiology of ADHD. The results also highlight strong overlap with the genetics of ADHD-related traits and health risk behaviors in the population, encouraging a dimensional view of ADHD as the extreme end of a continuum of symptoms.

URLs. LD-Hub, http://ldsc.broadinstitute.org/ldhub/; LD score regression, https://github.com/bulik/ldsc; Pre-computed European LD scores, https://data.broadinstitute.org/alkesgroup/LDSCORE/; PGC Ricopili GWA pipeline, https://github.com/Nealelab/ricopili; Credible set analysis, https://github.com/hailianghuang/FM-summary; FUMA, http://fuma.ctglab.nl.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41588-018-0269-7.

Received: 13 June 2017; Accepted: 28 September 2018;
Published online: 26 November 2018

References

1. Faraone, S. V. et al. Attention-deficit/hyperactivity disorder. *Nat. Rev. Dis. Primers*, 15020, https://doi.org/10.1038/nrdp.2015.20 (2015).
2. Dalsgaard, S., Leckman, J. F., Mortensen, P. B., Nielsen, H. S. & Simonsen, M. Effect of drugs on the risk of injuries in children with attention deficit hyperactivity disorder: a prospective cohort study. *Lancet Psychiatry* **2**, 702–709 (2015).
3. Chang, Z., Lichtenstein, P., D'Onofrio, B. M., Sjolander, A. & Larsson, H. Serious transport accidents in adults with attention-deficit/hyperactivity disorder and the effect of medication: a population-based study. *JAMA Psychiatry* **71**, 319–325 (2014).
4. Biederman, J. & Faraone, S. V. Attention-deficit hyperactivity disorder. *Lancet* **366**, 237–248 (2005).
5. Dalsgaard, S., Nielsen, H. S. & Simonsen, M. Consequences of ADHD medication use for children's outcomes. *J. Health Econ.* **37**, 137–151 (2014).
6. Dalsgaard, S., Mortensen, P. B., Frydenberg, M. & Thomsen, P. H. ADHD, stimulant treatment in childhood and subsequent substance abuse in adulthood - a naturalistic long-term follow-up study. *Addict. Behav.* **39**, 325–328 (2014).
7. Lichtenstein, P. & Larsson, H. Medication for attention deficit-hyperactivity disorder and criminality. *N. Engl. J. Med.* **368**, 776 (2013).
8. Barkley, R. A., Murphy, K. R. & Fischer, M. *ADHD in Adults: What the Science Says*. (Guilford Press, New York, 2007).
9. Furczyk, K. & Thome, J. Adult ADHD and suicide. *Atten. Defic. Hyperact. Disord.* **6**, 153–158 (2014).
10. Dalsgaard, S., Ostergaard, S. D., Leckman, J. F., Mortensen, P. B. & Pedersen, M. G. Mortality in children, adolescents, and adults with attention deficit hyperactivity disorder: a nationwide cohort study. *Lancet* **385**, 2190–2196 (2015).
11. Franke, B. et al. The genetics of attention deficit/hyperactivity disorder in adults, a review. *Mol. Psychiatry* **17**, 960–987 (2012).
12. Faraone, S. V. et al. Molecular genetics of attention-deficit/hyperactivity disorder. *Biol. Psychiatry* **57**, 1313–1323 (2005).
13. Burt, S. A. Rethinking environmental contributions to child and adolescent psychopathology: a meta-analysis of shared environmental influences. *Psychol. Bull.* **135**, 608–637 (2009).
14. Larsson, H., Anckarsater, H., Rastam, M., Chang, Z. & Lichtenstein, P. Childhood attention-deficit hyperactivity disorder as an extreme of a continuous trait: a quantitative genetic study of 8,500 twin pairs. *J. Child Psychol. Psychiatry* **53**, 73–80 (2012).
15. Christiansen, H. et al. Co-transmission of conduct problems with attention-deficit/hyperactivity disorder: familial evidence for a distinct disorder. *J. Neural Transm. (Vienna)* **115**, 163–175 (2008).

16. Kuntsi, J. et al. The separation of ADHD inattention and hyperactivity-impulsivity symptoms: pathways from genetic effects to cognitive impairments and symptoms. *J. Abnorm. Child Psychol.* **42**, 127–136 (2014).
17. Rommelse, N. N., Franke, B., Geurts, H. M., Hartman, C. A. & Buitelaar, J. K. Shared heritability of attention-deficit/hyperactivity disorder and autism spectrum disorder. *Eur. Child Adolesc. Psychiatry* **19**, 281–295 (2010).
18. Ghirardi, L. et al. The familial co-aggregation of ASD and ADHD: a register-based cohort study. *Mol. Psychiatry*. **23**, 257–262 (2018).
19. Larsson, H. et al. Risk of bipolar disorder and schizophrenia in relatives of people with attention-deficit hyperactivity disorder. *British J. Psychiatry* **203**, 103–106 (2013).
20. Faraone, S. V., Biederman, J. & Wozniak, J. Examining the comorbidity between attention deficit hyperactivity disorder and bipolar I disorder: a meta-analysis of family genetic studies. *Am. J. Psychiatry* **169**, 1256–1266 (2012).
21. Faraone, S. V. & Biederman, J. Do attention deficit hyperactivity disorder and major depression share familial risk factors? *J. Nerv. Ment. Dis.* **185**, 533–541 (1997).
22. Neale, B. M. et al. Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. *J. Am. Acad. Child Adolesc. Psychiatry* **49**, 884–897 (2010).
23. The Brainstorm Consortium. Analysis of shared heritability in common disorders of the brain. *Science* **360**, eaap8757 (2018).
24. Cross-Disorder Group of the Psychiatric Genomics Consortium. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* **45**, 984–994 (2013).
25. Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* **381**, 1371–1379 (2013).
26. Hamshere, M. L. et al. High loading of polygenic risk for ADHD in children with comorbid aggression. *Am. J. Psychiatry* **170**, 909–916 (2013).
27. Hamshere, M. L. et al. Shared polygenic contribution between childhood attention-deficit hyperactivity disorder and adult schizophrenia. *British J. Psychiatry* **203**, 107–111 (2013).
28. Groen-Blokhuis, M. M. et al. Attention-deficit/hyperactivity disorder polygenic risk scores predict attention problems in a population-based sample of children. *J. Am. Acad. Child Adolesc. Psychiatry* **53**, 1123–1129. e1126 (2014).
29. Martin, J., Hamshere, M. L., Stergiakouli, E., O'Donovan, M. C. & Thapar, A. Genetic risk for attention-deficit/hyperactivity disorder contributes to neurodevelopmental traits in the general population. *Biol. Psychiatry* **76**, 664–671 (2014).
30. Middeldorp, C. M. et al. A genome-wide association meta-analysis of attention-deficit/hyperactivity disorder symptoms in population-based pediatric cohorts. *J. Am. Acad. Child Adolesc. Psychiatry* **55**, 896–905.e896 (2016).
31. Yang, L. et al. Polygenic transmission and complex neuro developmental network for attention deficit hyperactivity disorder: genome-wide association study of both common and rare variants. *Am. J. Med. Genet. B Neuropsychiatr Genet.* **162B**, 419–430 (2013).
32. Zayats, T. et al. Genome-wide analysis of attention deficit hyperactivity disorder in Norway. *PLoS One* **10**, e0122501 (2015).
33. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
34. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
35. Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
36. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
37. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
38. Purcell, S. M. et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
39. Polanczyk, G., de Lima, M. S., Horta, B. L., Biederman, J. & Rohde, L. A. The worldwide prevalence of ADHD: a systematic review and metaregression analysis. *Am. J. Psychiatry* **164**, 942–948 (2007).
40. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
41. Zheng, J. et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272–279 (2017).
42. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
43. Wray, N. R. & Sullivan, P. F. et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).
44. Duncan, L. et al. Significant locus and metabolic genetic correlations revealed in genome-wide association study of anorexia nervosa. *Am. J. Psychiatry* **174**, 850–858 (2017).
45. Benyamin, B. et al. Childhood intelligence is heritable, highly polygenic and associated with FNBP1L. *Mol. Psychiatry* **19**, 253–258 (2014).
46. Okbay, A. et al. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat. Genet.* **48**, 624–633 (2016).
47. Rietveld, C. A. et al. GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* **340**, 1467–1471 (2013).
48. Rietveld, C. A. et al. Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proc. Natl Acad. Sci. USA* **111**, 13790–13794 (2014).
49. Davies, G. et al. Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N = 112 151). *Mol. Psychiatry* **21**, 758–767 (2016).
50. Teslovich, T. M. et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).
51. Morris, A. P. et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* **44**, 981–990 (2012).
52. Bradfield, J. P. et al. A genome-wide association meta-analysis identifies new childhood obesity loci. *Nat. Genet.* **44**, 526–531 (2012).
53. Berndt, S. I. et al. Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nat. Genet.* **45**, 501–512 (2013).
54. Speliotes, E. K. et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42**, 937–948 (2010).
55. Shungin, D. et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187–196 (2015).
56. Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat. Genet.* **42**, 441–447 (2010).
57. Patel, Y. M. et al. Novel Association of genetic markers affecting CYP2A6 activity and lung cancer risk. *Cancer Res.* **76**, 5768–5776 (2016).
58. Wang, Y. et al. Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer. *Nat. Genet.* **46**, 736–741 (2014).
59. Barban, N. et al. Genome-wide analysis identifies 12 loci influencing human reproductive behavior. *Nat. Genet.* **48**, 1462–1472 (2016).
60. Hammerschlag, A. R. et al. Genome-wide association analysis of insomnia complaints identifies risk genes and genetic overlap with psychiatric and metabolic traits. *Nat. Genet.* **49**, 1584–1592 (2017).
61. Pilling, L. C. et al. Human longevity is influenced by many genetic variants: evidence from 75,000 UK Biobank participants. *Aging* **8**, 547–560 (2016).
62. Hawi, Z. et al. The molecular genetic architecture of attention deficit hyperactivity disorder. *Mol. Psychiatry* **20**, 289–297 (2015).
63. Sia, G. M., Clem, R. L. & Huganir, R. L. The human language-associated gene SRPX2 regulates synapse formation and vocalization in mice. *Science* **342**, 987–991 (2013).
64. Tsui, D., Vessey, J. P., Tomita, H., Kaplan, D. R. & Miller, F. D. FoxP2 regulates neurogenesis during embryonic cortical development. *J. Neurosci.* **33**, 244–258 (2013).
65. Schreiweis, C. et al. Humanized Foxp2 accelerates learning by enhancing transitions from declarative to procedural performance. *Proc. Natl. Acad. Sci. USA* **111**, 14253–14258 (2014).
66. Jensen, C. M. & Steinhausen, H. C. Comorbid mental disorders in children and adolescents with attention-deficit/hyperactivity disorder in a large nationwide study. *Atten. Defic. Hyperact. Disord* **7**, 27–38 (2015).
67. Larson, K., Russ, S. A., Kahn, R. S. & Halfon, N. Patterns of comorbidity, functioning, and service use for US children with ADHD, 2007. *Pediatrics* **127**, 462–470 (2011).
68. Peyre, H. et al. Relationship between early language skills and the development of inattention/hyperactivity symptoms during the preschool period: Results of the EDEN mother-child cohort. *BMC Psychiatry* **16**, 380 (2016).
69. Breiderhoff, T. et al. Sortilin-related receptor SORCS3 is a postsynaptic modulator of synaptic depression and fear extinction. *PLoS One* **8**, e75006 (2013).
70. Hyde, C. L. et al. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nat. Genet.* **48**, 1031–1036 (2016).
71. Caunt, C. J. & Keyse, S. M. Dual-specificity MAP kinase phosphatases (MKPs): shaping the outcome of MAP kinase signalling. *FEBS. J.* **280**, 489–504 (2013).

72. Mortensen, O. V. MKP3 eliminates depolarization-dependent neurotransmitter release through downregulation of L-type calcium channel Cav1.2 expression. *Cell Calcium* **53**, 224–230 (2013).
73. Mortensen, O. V., Larsen, M. B., Prasad, B. M. & Amara, S. G. Genetic complementation screen identifies a mitogen-activated protein kinase phosphatase, MKP3, as a regulator of dopamine transporter trafficking. *Mol. Biol. Cell.* **19**, 2818–2829 (2008).
74. Volkow, N. D., Fowler, J. S., Wang, G., Ding, Y. & Gatley, S. J. Mechanism of action of methylphenidate: insights from PET imaging studies. *J. Atten. Disord* **6**(Suppl 1), S31–S43 (2002).
75. Volkow, N. D. et al. Therapeutic doses of oral methylphenidate significantly increase extracellular dopamine in the human brain. *J. Neurosci.* **21**, RC121 (2001).
76. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
77. Qu, X. et al. Identification, characterization, and functional study of the two novel human members of the semaphorin gene family. *J. Biol. Chem.* **277**, 35574–35585 (2002).
78. Okbay, A. et al. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539–542 (2016).
79. Zhernakova, D. V. et al. Identification of context-dependent expression quantitative trait loci in whole blood. *Nat. Genet.* **49**, 139–145 (2017).
80. Hu, H. et al. ST3GAL3 mutations impair the development of higher cognitive functions. *Am. J. Hum. Genet.* **89**, 407–414 (2011).
81. Oliver, P. L. et al. Disruption of Visc-2, a brain-expressed conserved long noncoding rna, does not elicit an overt anatomical or behavioral phenotype. *Cereb. Cortex* **25**, 3572–3585 (2015).
82. Sobreira, N., Walsh, M. F., Batista, D. & Wang, T. Interstitial deletion 5q14.3-q21 associated with iris coloboma, hearing loss, dental anomaly, moderate intellectual disability, and attention deficit and hyperactivity disorder. *Am. J. Med. Genet. A.* **149A**, 2581–2583 (2009).
83. Le Meur, N. et al. MEF2C haploinsufficiency caused by either microdeletion of the 5q14.3 region or mutation is responsible for severe mental retardation with stereotypic movements, epilepsy and/or cerebral malformations. *J. Med. Genet.* **47**, 22–29 (2010).
84. Novara, F. et al. Refining the phenotype associated with MEF2C haploinsufficiency. *Clin. Genet.* **78**, 471–477 (2010).
85. Lambert, J. C. et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45**, 1452–1458 (2013).
86. Gene Ontology Consortium. Gene Ontology Consortium: going forward. *Nucleic Acids Res.* **43**, D1049–D1056 (2015).
87. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).
88. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
89. Vernes, S. C. et al. Foxp2 regulates gene networks implicated in neurite outgrowth in the developing brain. *PLoS. Genet.* **7**, e1002145 (2011).
90. Spiteri, E. et al. Identification of the transcriptional targets of FOXP2, a gene linked to speech and language, in developing human brain. *Am. J. Hum. Genet.* **81**, 1144–1157 (2007).
91. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
92. Ebejer, J. L. et al. Genome-wide association study of inattention and hyperactivity-impulsivity measured as quantitative traits. *Twin. Res. Hum. Genet.* **16**, 560–574 (2013).
93. Grove, J. et al. Common risk variants identified in autism spectrum disorder. *bioRxiv*. https://doi.org/10.1101/224774 (2017).
94. Lindblad-Toh, K. et al. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**, 476–482 (2011).
95. Flory, K. et al. Childhood ADHD predicts risky sexual behavior in young adulthood. *J Clin Child Adolesc. Psychol.* **35**, 571–577 (2006).
96. Marsh, L. E., Norvilitis, J. M., Ingersoll, T. S. & Li, B. ADHD symptomatology, fear of intimacy, and sexual anxiety and behavior among college students in China and the United States. *J. Atten. Disord.* **19**, 211–221 (2015).
97. Hosain, G. M., Berenson, A. B., Tennen, H., Bauer, L. O. & Wu, Z. H. Attention deficit hyperactivity symptoms and risky sexual behavior in young adult women. *J. Womens Health (Larchmt)* **21**, 463–468 (2012).
98. Chudal, R. et al. Parental age and the risk of attention-deficit/hyperactivity disorder: a nationwide, population-based cohort study. *J. Am. Acad. Child Adolesc. Psychiatry* **54**, 487–494.e481 (2015).
99. Chang, Z. et al. Maternal age at childbirth and risk for ADHD in offspring: a population-based cohort study. *Int. J. Epidemiol.* **43**, 1815–1824 (2014).
100. Ostergaard, S. D., Dalsgaard, S., Faraone, S. V., Munk-Olsen, T. & Laursen, T. M. Teenage parenthood and birth rates for individuals with and without attention-deficit/hyperactivity disorder: a nationwide cohort study. *J. Am. Acad. Child Adolesc. Psychiatry* **56**, 578–584.e573 (2017).
101. Barbaresi, W. J., Katusic, S. K., Colligan, R. C., Weaver, A. L. & Jacobsen, S. J. Long-term school outcomes for children with attention-deficit/hyperactivity disorder: a population-based perspective. *J. Dev. Behav. Pediatr.* **28**, 265–273 (2007).
102. Faraone, S. V. et al. Intellectual performance and school failure in children with attention deficit hyperactivity disorder and in their siblings. *J. Abnorm. Psychol.* **102**, 616–623 (1993).
103. Sniekers, S. et al. Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nat. Genet.* **49**, 1107–1112 (2017).
104. Kong, A. et al. Selection against variants in the genome associated with educational attainment. *Proc. Natl. Acad. Sci. USA* **114**, E727–E732 (2017).
105. Lee, S. S., Humphreys, K. L., Flory, K., Liu, R. & Glass, K. Prospective association of childhood attention-deficit/hyperactivity disorder (ADHD) and substance use and abuse/dependence: a meta-analytic review. *Clin. Psychol. Rev.* **31**, 328–341 (2011).
106. Halfon, N., Larson, K. & Slusser, W. Associations between obesity and comorbid mental health, developmental, and physical health conditions in a nationally representative sample of US children aged 10 to 17. *Acad. Pediatr.* **13**, 6–13 (2013).
107. Chen, A. Y., Kim, S. E., Houtrow, A. J. & Newacheck, P. W. Prevalence of obesity among children with chronic conditions. *Obesity (Silver Spring)* **18**, 210–213 (2010).
108. Cortese, S. et al. Association between ADHD and obesity: A systematic review and meta-analysis. *Am. J. Psychiatry* **173**, 34–43 (2016).
109. Owens, J. A. A clinical overview of sleep and attention-deficit/hyperactivity disorder in children and adolescents. *J. Can. Acad. Child Adolesc. Psychiatry* **18**, 92–102 (2009).
110. Lubke, G. H., Hudziak, J. J., Derks, E. M., van Bijsterveldt, T. C. & Boomsma, D. I. Maternal ratings of attention problems in ADHD: evidence for the existence of a continuum. *J. Am. Acad. Child Adolesc. Psychiatry* **48**, 1085–1093 (2009).
111. Cortese, S., Comencini, E., Vincenzi, B., Speranza, M. & Angriman, M. Attention-deficit/hyperactivity disorder and impairment in executive functions: a barrier to weight loss in individuals with obesity? *BMC Psychiatry* **13**, 286 (2013).
112. Ortal, S. et al. The role of different aspects of impulsivity as independent risk factors for substance use disorders in patients with ADHD: a review. *Curr. Drug Abuse Rev.* **8**, 119–133 (2015).

Acknowledgements

The iPSYCH team acknowledges funding from the Lundbeck Foundation (grant no. R102-A9118 and R155-2014-1724), the Stanley Medical Research Institute, the European Research Council (project 294838), the European Community (EC) Horizon 2020 Programme (grant 667302 (CoCA)), from EC Seventh Framework Programme (grant 602805 (Aggressotype)), the Novo Nordisk Foundation for supporting the Danish National Biobank resource and grants from Aarhus and Copenhagen Universities and University Hospitals, including support to the iSEQ Center, the GenomeDK HPC facility, and the CIRRAU Center.

The Broad Institute and Massachusetts General Hospital investigators would like to acknowledge support from the Stanley Medical Research Institute and NIH grants: 5U01MH094432-04(PI: Daly), 1R01MH094469 (PI: Neale), 1R01MH107649-01 (PI: Neale), 1R01MH109539-01 (PI: Daly).

We thank T. Lehner, A. Addington and G. Senthil for their support in the Psychiatric Genomics Consortium.

S.V.F. is supported by the K.G. Jebsen Centre for Research on Neuropsychiatric Disorders, University of Bergen, Norway, the EC's Seventh Framework Programme (grant 602805), the EC's Horizon 2020 (grant 667302) and NIMH grants 5R01MH101519 and U01 MH109536-01.

J.M. was supported by the Wellcome Trust (grant 106047).

B.F.'s research is supported by funding from a personal Vici grant of the Netherlands Organisation for Scientific Research (NWO; grant 016-130-669, to B.F.), the EC's Seventh Framework Programme (grant 602805 (Aggressotype), 602450 (IMAGEMEND), and 278948 (TACTICS)), and from the EC's Horizon 2020 Programme (grant643051 (MiND) and 667302 (CoCA)). Additionally, this work was supported by the European College of Neuropsychopharmacology (ECNP Network 'ADHD across the Lifespan').

J.H. is supported by grants from Stiftelsen K.G. Jebsen, University of Bergen and The Research Council of Norway.

B.C. received financial support for this research from the Spanish 'Ministerio de Economía y Competitividad' (SAF2015-68341-R) and 'Generalitat de Catalunya/AGAUR' (2017-SGR-738). B.B., A.R. and collaborators received funding from the EC's Seventh Framework Programme (grant 602805, Aggressotype), the EC's H2020 Programme (grants 667302, CoCA, and 402003, MiND), the ECNP network 'ADHD across the lifespan' and DFG CRC 1193, subproject Z03.

O.A.A. is supported by the Research Council of Norway (grants: 223273, 248778, 213694, 249711), and KG Jebsen Stiftelsen.

A.T. received ADHD funding from the Wellcome Trust, Medical Research Council (MRC UK), Action Medical Research.

We thank the customers of 23andMe who answered surveys, as well as the employees of 23andMe who together made this research possible. The QIMR studies were supported by funding from the Australian National Health and Medical Research Council (grant numbers: 241944, 339462, 389927, 389875, 389891, 389892, 389938, 443036, 442915, 442981, 496739, 552485, and 552498, and, most recently, 1049894) and the Australian Research Council (grant numbers: A7960034, A79906588, A79801419, DP0212016, and DP0343921). SEM is supported by an NHMRC fellowship (1103623).

Additional acknowledgements can be found in the Supplementary Note.

Author contributions

Analysis: D.D., R.K.W., J. Martin, M.M., T.D.A., C.C., N.E., M.G., K.L.G., M.E.H., D.P.H., H. Huang, J.B.M., A.R.M., J.P., D.S.P., T.P., S.R., E.B.R., F.K.S., H.S., P.T., G.B.W., H.W., D.I.B., D.G., C.M., P.R., P.F.S., J.Y.T., S.E.M., K.S., A.D.B. and B.M.N. supervised and coordinated analyses. **Sample and/or data provider and processing:** D.D., R.K.W., J. Martin, M.M., E.A., G.B., R.B., J.B.-G., M.B.-H., F.C., K.C., A.D., N.E., J.I.G., J. Grove, O.O.G., C.S.H., M.V.H., J.B.M., N.G.M., J. Moran, C.B.P., M.G.P., J.B.P., S.R., C.S., M.J.W., O.A.A., P.A., C.L.B., D.I.B., B.C., S.D., B.F., J. Gelernter, H. Hakonarson, J.H., H.R.K., J.K., K.L., K.-P.L., C.M., A.R., L.A.R., R.S., P.S., E.J.S.S.-B., A.T., J.Y.T., I.D.W., S.E.M., D.M.H., O.M., P.B.M., A.D.B., ADHD Working Group of the Psychiatric Genomics Consortium, Early Lifecourse & Genetic Epidemiology (EAGLE) Consortium, 23andMe Research Team. **Core PI group:** S.E.M., K.S., M.N., D.M.H., T.W., O.M., P.B.M., M.J.D., S.V.F., A.D.B., B.M.N. **Core writing group:** D.D., R.K.W., J. Martin, S.V.F., A.D.B., B.M.N. **Direction of study:** A.D.B., S.V.F., B.M.N. All authors contributed with critical revision of the manuscript.

Competing interests

In the past year, S.V.F. received income, potential income, travel expenses, continuing education support and/or research support from Lundbeck, Rhodes, Arbor, KenPharm, Ironshore, Shire, Akili Interactive Labs, CogCubed, Alcobra, VAYA, Sunovion, Genomind and Neurolifesciences. With his institution, he has US patent US20130217707 A1 for the use of sodium–hydrogen exchange inhibitors in the treatment of ADHD. In previous years, he received support from: Shire, Neurovance, Alcobra, Otsuka, McNeil, Janssen, Novartis, Pfizer and Eli Lilly. S.V.F. receives royalties from books published by Guilford Press: *Straight Talk about Your Child's Mental Health*; Oxford University Press: *Schizophrenia: The Facts*; and Elsevier: *ADHD: Non-Pharmacologic Interventions*. He is principal investigator of www.adhdinadults.com.

B.M.N. is a member of Deep Genomics Scientific Advisory Board and has received travel expenses from Illumina. He also serves as a consultant for Avanir and Trigeminal solutions.

O.O.G., G.B.W., H.S. and K.S. are employees of deCODE genetics/Amgen.

N.E., J.Y.T., and the 23andMe Research Team are employees of 23andMe, Inc. and hold stock or stock options in 23andMe.

L.A.R. has received honoraria, has been on the speakers' bureau/advisory board and/or has acted as a consultant for Eli-Lilly, Janssen-Cilag, Novartis, Medice and Shire in the past three years. He receives authorship royalties from Oxford Press and ArtMed. He also received a travel award from Shire for taking part in the 2015 WFADHD meeting. The ADHD and Juvenile Bipolar Disorder Outpatient Programs unrestricted educational and research support from the following pharmaceutical companies in the past three years: Eli-Lilly, Janssen-Cilag, Novartis and Shire. Over the past three years E.J.S.-B. has received speaker fees, consultancy, research funding and conference support from Shire Pharma and speaker fees from Janssen-Cilag. He has received consultancy fees from Neurotech solutions, Aarhus University, Copenhagen University and Berhanderling, Skolerne, Copenhagen, KU Leuven and book royalties from OUP and Jessica Kingsley. He is the editor-in-chief of the *Journal of Child Psychology and Psychiatry*, for which his university receives financial support. B.F. has received educational speaking fees from Merz and Shire.

R.S. has equity in and is on the advisory board of Ironshore Pharmaceuticals. A.R. has received a research grant from Medice and speaker's honorarium from Medice and Servier. J.H. has received speaker fees from Shire, Lilly and Novartis. H.R.K. has been an advisory board member, consultant, or CME speaker for Alkermes, Indivior, and Lundbeck. He is also a member of the American Society of Clinical Psychopharmacology's Alcohol Clinical Trials Initiative, which was supported in the last three years by AbbVie, Alkermes, Ethypharm, Indivior, Lilly, Lundbeck, Otsuka, Pfizer, Arbor, and Amygdala Neurosciences. H.R.K. and J.G. are named as inventors on PCT patent application #15/878,640 entitled: "Genotype-guided dosing of opioid agonists," filed January 24, 2018. P.A. received honoraria paid to King's College London by Shire, Flynn Pharma, Lilly, Janssen, Novartis and Lunbeck for research, speaker fees, education events, advisory board membership or consultancy. O.A.A. has received speaker fees from Lundbeck and Sunovion. J.K. has received speaker's honorarium from Medice; all funds are received by King's College London and used for studies of ADHD. T.W. has acted as lecturer and scientific advisor to H. Lundbeck A/S.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-018-0269-7.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to S.V.F. or A.D.B. or B.M.N.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

[1]The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Aarhus, Denmark. [2]Centre for Integrative Sequencing, iSEQ, Aarhus University, Aarhus, Denmark. [3]Department of Biomedicine - Human Genetics, Aarhus University, Aarhus, Denmark. [4]Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. [5]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [6]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [7]MRC Centre for Neuropsychiatric Genetics & Genomics, School of Medicine, Cardiff University, Cardiff, UK. [8]Centre for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden. [9]Stockholm Health Care Services, Stockholm County Council, Stockholm, Sweden. [10]Department of Psychiatry, Psychosomatics and Psychotherapy, University of Wuerzburg, Wuerzburg, Germany. [11]National Centre for Register-Based Research, Aarhus University, Aarhus, Denmark. [12]Centre for Integrated Register-based Research, Aarhus University, Aarhus, Denmark. [13]Department of Child and Adolescent Psychiatry, National University Hospital, Reykjavik, Iceland. [14]Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark. [15]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [16]23andMe, Inc, Mountain View, CA, USA. [17]Program in Neurogenetics, Department of Neurology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. [18]Center for Autism Research and Treatment and Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, Los Angeles, CA, USA. [19]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. [20]Department of Psychiatry, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, Los Angeles, CA, USA. [21]QIMR Berghofer Medical Research Institute, Brisbane, Australia. [22]Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark. [23]deCODE genetics/Amgen, Reykjavík, Iceland. [24]Faculty of Medicine, University of Iceland, Reykjavík, Iceland. [25]Institute of Biological Psychiatry, MHC Sct. Hans, Mental Health Services Copenhagen, Roskilde, Denmark. [26]Genomics plc, Oxford, UK. [27]Department of Psychiatry and Psychotherapy, Charité - Universitätsmedizin, Berlin, Germany. [28]Department of Epidemiology, Harvard Chan School of Public Health, Boston, MA, USA. [29]Queensland Brain Institute, University of Queensland, Brisbane, Australia. [30]A list of members and affiliations appears at the end of the paper. [31]NORMENT KG Jebsen Centre for Psychosis Research, Division of Mental Health and Addiction, University of Oslo and Oslo University Hospital, Oslo, Norway. [32]Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK. [33]Psychiatry, Neurosciences and Mental Health, The Hospital for Sick Children, University of Toronto, Toronto, Canada. [34]Department of Biological Psychology, Neuroscience Campus Amsterdam, VU University, Amsterdam, The Netherlands. [35]EMGO Institute for Health and Care Research, Amsterdam, The Netherlands. [36]Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona, Barcelona, Catalonia, Spain. [37]Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Instituto de Salud Carlos III, Madrid, Spain. [38]Institut de Biomedicina de la Universitat de Barcelona (IBUB), Barcelona, Catalonia, Spain. [39]Institut de Recerca Sant Joan de Déu (IRSJD), Esplugues de Llobregat, Barcelona, Catalonia, Spain. [40]Departments of Human Genetics (855) and Psychiatry, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Centre, Nijmegen, The Netherlands. [41]Department of Psychiatry, Genetics, and Neuroscience, Yale University School of Medicine, New Haven, CT, USA. [42]Veterans Affairs Connecticut Healthcare Center, West Haven, CT, USA. [43]The Center for Applied Genomics, The Children´s Hospital of Philadelphia, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [44]K.G. Jebsen Centre for

Neuropsychiatric Disorders, Department of Biomedicine, University of Bergen, Bergen, Norway. [45]Haukeland University Hospital, Bergen, Norway. [46]Department of Psychiatry, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [47]Veterans Integrated Service Network (VISN4) Mental Illness Research, Education, and Clinical Center (MIRECC), Crescenz VA Medical Center, Philadephia, PA, USA. [48]School of Psychology, Cardiff University, Cardiff, UK. [49]Division of Molecular Psychiatry, Center of Mental Health, University of Wuerzburg, Wuerzburg, Germany. [50]Department of Neuroscience, School for Mental Health and Neuroscience (MHENS), Maastricht University, Maastricht, The Netherlands. [51]Laboratory of Psychiatric Neurobiology, Institute of Molecular Medicine, I.M. Sechenov First Moscow State Medical University, Moscow, Russia. [52]Child Health Research Centre, University of Queensland, Brisbane, Australia. [53]Child and Youth Mental Health Service, Children's Health Queensland Hospital and Health Service, Brisbane, Australia. [54]Department of Psychiatry, Psychosomatic Medicine and Psychotherapy, University Hospital Frankfurt, Frankfurt am Main, Germany. [55]Department of Psychiatry, Faculty of Medicine, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil. [56]ADHD Outpatient Clinic, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil. [57]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [58]Institute for Genomics and Multiscale Biology, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [59]Friedman Brain Institute, Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [60]Mental Illness Research Education and Clinical Center (MIRECC), James J. Peters VA Medical Center, Bronx, New York, USA. [61]Institute of Psychiatry, Psychology & Neuroscience, Kings College, London, UK. [62]Departments of Genetics and Psychiatry, University of North Carolina, Chapel Hill, NC, USA. [63]Department of Psychology, Emory University, Atlanta, GA, USA. [64]Mental Health Services in the Capital Region of Denmark, Mental Health Center Copenhagen, University of Copenhagen, Copenhagen, Denmark. [65]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark. [66]Psychosis Research Unit, Aarhus University Hospital, Risskov, Denmark. [67]Institute for Molecular Medicine Finland (FIMM), Helsinki, Finland. [68]Departments of Psychiatry and Neuroscience and Physiology, SUNY Upstate Medical University, Syracuse, NY, USA. [69]These authors contributed equally: Ditte Demontis, Raymond K. Walters. [70]These authors jointly supervised this work: Stephen V. Faraone, Anders D. Børglum, Benjamin M. Neale.
*e-mail: sfaraone@childpsychresearch.org; anders@biomed.au.dk; bneale@broadinstitute.org

ADHD Working Group of the Psychiatric Genomics Consortium (PGC)

Özgür Albayrak[71,72], Richard J. L. Anney[73], Maria Jesús Arranz[74], Tobias J. Banaschewski[75], Claiton Bau[56,76], Joseph Biederman[77,78], Jan K. Buitelaar[79,80], Miguel Casas[81,82,83,84], Alice Charach[85], Jennifer Crosbie[85], Astrid Dempfle[86], Alysa E. Doyle[87,88], Richard P. Ebstein[89], Josephine Elia[90,91], Christine Freitag[92], Manuel Föcker[71], Michael Gill[93], Eugenio Grevet[55,56], Ziarih Hawi[94], Johannes Hebebrand[71], Beate Herpertz-Dahlmann[95], Amaia Hervas[74], Anke Hinney[71], Sarah Hohmann[75], Peter Holmans[73], Mara Hutz[76], Abel Ickowitz[85], Stefan Johansson[96], Lindsey Kent[97], Sarah Kittel-Schneider[98], Nanda Lambregts-Rommelse[99], Gerd Lehmkuhl[100], Sandra K. Loo[101], James J. McGough[102], Jobst Meyer[103], Eric Mick[104], Frank Middleton[105], Ana Miranda[106], Nina Roth Mota[56,107], Fernando Mulas[108], Aisling Mulligan[109], Freimer Nelson[110], T. Trang Nguyen[111], Robert D. Oades[112], Michael C. O'Donovan[73], Michael J. Owen[73], Haukur Palmason[113], Josep Antoni Ramos-Quiroga[37,83,114,115], Tobias J. Renner[116,117], Marta Ribasés[37,83,114], Marcella Rietschel[118], Olga Rivero[49], Jasmin Romanos[119], Marcel Romanos[120], Aribert Rothenberger[121], Herbert Royers[122], Christina Sánchez-Mora[37,83,114], André Scherag[123,124], Benno G. Schimmelmann[125], Helmut Schäfer[111], Joseph Sergeant[126], Judith Sinzig[100,127], Susan L. Smalley[128], Hans-Christoph Steinhausen[129,130,131], Margaret Thompson[132], Alexandre Todorov[133], Alejandro Arias Vasquez[134], Susanne Walitza[119,135], Yufeng Wang[136], Andreas Warnke[119], Nigel Williams[137], Stephanie H. Witt[118], Li Yang[136], Tetyana Zayats[44,138] and Yanli Zhang-James[105]

Early Lifecourse & Genetic Epidemiology (EAGLE) Consortium

George Davey Smith[139], Gareth E. Davies[34,140], Erik A. Ehli[140], David M. Evans[139,141], Iryna O. Fedko[34], Corina U. Greven[80,142,143], Maria M. Groen-Blokhuis[144], Monica Guxens[37,145,146,147], Anke R. Hammerschlag[148], Catharina A. Hartman[149], Joachim Heinrich[150,151], Jouke Jan Hottenga[152], James Hudziak[153,154,155,156], Astanand Jugessur[157,158], John P. Kemp[139,141], Eva Krapohl[143], Mario Murcia[37,159], Ronny Myhre[160], Ilja M. Nolte[161], Dale R. Nyholt[162], Johan Ormel[149], Klaasjan G. Ouwens[34],

Irene Pappa[147,163], Craig E. Pennell[164], Robert Plomin[143], Susan Ring[139,165], Marie Standl[150], Evie Stergiakouli[139,141], Beate St Pourcain[139,166], Camilla Stoltenberg[167], Jordi Sunyer[146,168,169], Elisabeth Thiering[150,170], Henning Tiemeier[155], Carla M.T. Tiesler[150,170], Nicholas J. Timpson[139], Maciej Trzaskowski[29], Peter Johannes van der Most[161], Natalia Vilor-Tejedor[145,146,168], Carol A. Wang[164], Andrew J.O. Whitehouse[171] and Huiying Zhao[162]

23andMe Research Team

Michelle Agee[16], Babak Alipanahi[16], Adam Auton[16], Robert K. Bell[16], Katarzyna Bryc[16], Sarah L. Elson[16], Pierre Fontanillas[16], Nicholas A. Furlotte[16], David A. Hinds[16], Bethann S. Hromatka[16], Karen E. Huber[16], Aaron Kleinman[16], Nadia K. Litterman[16], Matthew H. McIntyre[16], Joanna L. Mountain[16], Carrie A. M. Northover[16], Steven J. Pitts[16], J. Fah Sathirapongsasuti[16], Olga V. Sazonova[16], Janie F. Shelton[16], Suyash Shringarpure[16], Chao Tian[16], Vladimir Vacic[16] and Catherine H. Wilson[16]

[71]Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University Hospital Essen, University of Duisburg-Essen, Essen, Germany. [72]Department of Psychosomatic Medicine and Psychotherapy, Hannover Medical School (MHH), Hannover, Germany. [73]MRC Centre for Neuropsychiatric Genetics and Genomics, Department of Psychological Medicine and Neurology, School of Medicine, Cardiff University, Cardiff, Wales, UK. [74]University Hospital Mutua Terrassa, Barcelona, Spain. [75]Department of Child and Adolescent Psychiatry, Central Institute of Mental Health and Mannheim Medical Faculty, University of Heidelberg, Heidelberg, Germany. [76]Department of Genetics, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil. [77]Pediatric Psychopharmacology Unit, Massachusetts General Hospital, Boston, MA, USA. [78]Department of Psychiatry, Harvard Medical School, Boston, MA, USA. [79]Department of Cognitive Neuroscience, Donders Institute for Brain, Cognition and Behavior, Radboud University Medical Centre, Nijmegen, The Netherlands. [80]Karakter Child and Adolescent Psychiatry University Center, Nijmegen, The Netherlands. [81]Universitat Autònoma de Barcelona, Barcelona, Spain. [82]Programa Corporatiu "Neurodevelopment Disorders along Life Span", Institut Català de la Salut, Barcelona, Spain. [83]Department of Psychiatry, Hospital Universitari Vall d'Hebron, Barcelona, Spain. [84]Clinica Galatea y PAIMM, Mental Health Program for Impaired Physicians, Barcelona, Spain. [85]The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada. [86]Institute of Medical Informatics and Statistics, Kiel University, Kiel, Germany. [87]Massachusetts General Hospital, Boston, MA, USA. [88]Harvard Medical School, Boston, MA, USA. [89]National University of Singapore, Singapore, Singapore. [90]Department of Pediatrics, Nemours A.I. duPont Hospital for Children, Wilmington, DE, USA. [91]Department of Psychiatry, Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, PA, USA. [92]Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University Hospital Frankfurt, Goethe University, Frankfurt am Main, Germany. [93]Department of Psychiatry, Trinity College Dublin, Trinity Centre for Health Sciences, St. James's Hospital, Dublin, Ireland. [94]School of Psychological Sciences and Monash Institute for Cognitive and Clinical Neurosciences, Monash University, Melbourne, Australia. [95]Department of Child & Adolescent Psychiatry & Psychosomatic Medicine of University Clinics, RWTH Aachen, Aachen, Germany. [96]K.G.Jebsen Centre for Psychiatric Disorders, Department of Clinical Science, University of Bergen, Bergen, Norway. [97]University of St Andrews, St Andrews, UK. [98]Department of Psychiatry, Psychosomatic Medicine and Psychotherapy, University Hospital, Frankfurt, Germany. [99]Karakter Child and Adolescent Psychiatry University Center and department of Psychiatry, Donders Institute for Brain, Cognition and Behavior, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands. [100]Department of Child and Adolescent Psychiatry, University of Cologne, Cologne, Germany. [101]Department of Psychiatry, University of California, Los Angeles, Los Angeles, CA, USA. [102]Semel Institute for Neuroscience & Human Behavior, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, CA, USA. [103]Institute of Psychobiology, Department of Neurobehavioral Genetics, University of Trier, Trier, Germany. [104]Quantitative Health Sciences University of Massachusetts Medical School, Worcester, MA, USA. [105]Department of Psychiatry, SUNY Upstate Medical University, Syracuse, NY, USA. [106]Department of Developmental and Educational Psychology, University of Valencia, Valencia, Spain. [107]Department of Human Genetics, Radboud University Medical Center, Nijmegen, Netherlands. [108]Instituto Valenciano de Neurologia Pediatrica (INVANEP), Valencia, Spain. [109]Senior Lecturer in Child and Adolescent Psychiatry, University College Dublin, Dublin, Ireland. [110]Center for Neurobehavioral Genetics, Semel Institute for Neuroscience & Human Behavior, University of California at Los Angeles, Los Angeles, CA, USA. [111]University of Marburg, Marburg, Germany. [112]Clinic for Child and Adolescent Psychiatry and Psychotherapy, University of Duisburg-Essen, Essen, Germany. [113]Landspitali National University Hospital, Reykjavik, Iceland. [114]Psychiatric Genetics Unit, Group of Psychiatry, Mental Health and Addiction, Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona, Barcelona, Spain. [115]Department of Psychiatry and Legal Medicine, Universitat Autònoma de Barcelona, Barcelona, Spain. [116]Department of Child and Adolescent Psychiatry, Universitätsklinikum Tübingen, Tübingen, Germany. [117]Division of Molecular Psychiatry, ADHD Clinical Research Unit, Department of Psychiatry, Psychosomatics and Psychotherapy, University of Wuerzburg, Wuerzburg, Germany. [118]Central Institute of Mental Health, Department of Genetic Epidemiology in Psychiatry, Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany. [119]Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University of Wuerzburg, Wuerzburg, Germany. [120]University Hospital of Würzburg, Center of Mental Health, Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, Wuerzburg, Germany. [121]Child and Adolescent Psychiatry/Psychotherapy, University Medical Center, Goettingen, Germany. [122]Ghent University, Dunantlaan, Ghent, Belgium. [123]Institute for Medical Informatics, Biometry and Epidemiology (IMIBE), University of Duisburg-Essen, Essen, Germany. [124]Clinical Epidemiology, Integrated Research and Treatment Center, Center for Sepsis Control and Care (CSCC), Jena University Hospital, Jena, Germany. [125]University Hospital of Child- and Adolescent Psychiatry, University of Bern, Bern, Switzerland. [126]Vrije Universiteit, De Boelelaan, Amsterdam, The Netherlands. [127]Department of Child and Adolescent Psychiatry and Psychotherapy, LVR – Clinic Bonn, Bonn, Germany. [128]University of California Los Angeles, Los Angeles, CA, USA. [129]University of Zurich, Zurich, Switzerland. [130]Aalborg University, Aalborg, Denmark. [131]University of Basel, Basel, Switzerland. [132]University of Southampton, Southampton, UK. [133]Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA. [134]Department of Psychiatry & Human Genetics, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, The Netherlands. [135]Department of Child and Adolescent Psychiatry, University of Zurich, Zurich, Switzerland. [136]Peking University Institute of Mental Health, Beijing Shi, China. [137]Cardiff University, Medical Research Council Center for Neuropsychiatric Genetics & Genomics, Institute of Psychology, Medicine & Clinical Neuroscience, Cardiff, UK. [138]Analytic

Translational Genetics Unit, Massachusetts General Hospital Harvard Medical School, Boston, MA, USA. [139]Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol, UK. [140]Avera Institute for Human Genetics, Sioux Falls, SD, USA. [141]University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, Queensland, Australia. [142]Radboud University Medical Center, Donders Institute for Brain, Cognition and Behaviour, Department of Cognitive Neuroscience, Nijmegen, The Netherlands. [143]Medical Research Council Social, Genetic & Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK. [144]GGZ inGeest, Amsterdam, The Netherlands. [145]ISGlobal - Centre for Research in Environmental Epidemiology (CREAL), Barcelona, Spain. [146]Universitat Pompeu Fabra (UPF), Barcelona, Spain. [147]Department of Child and Adolescent Psychiatry/Psychology, Erasmus University Medical Centre–Sophia Children's Hospital, Rotterdam, The Netherlands. [148]Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, VU University Amsterdam, Amsterdam, The Netherlands. [149]Department of Psychiatry, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands. [150]Institute of Epidemiology I, Helmholtz Zentrum München – German Research Centre for Environmental Health, Neuherberg, Germany. [151]Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine, Inner City Clinic, University Hospital of Munich (LMU), Munich, Germany. [152]Biological Psychology, Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. [153]Vermont Center for Children Youth and Families and University of Vermont Medical Center, University of Vermont, Burlington, VT, USA. [154]Child Psychiatry, School of Medicine, Washington University, St. Louis, MO, USA. [155]Erasmus University Medical Centre–Sophia Children's Hospital, Rotterdam, The Netherlands. [156]Geisel School of Medicine, Dartmouth, Hanover, NH, USA. [157]Department of Genetic Research and Bioinformatics, Norwegian Institute of Public Health, Oslo, Norway. [158]Department of Global Public Health and Primary Care, University of Bergen, Bergen, Norway. [159]Epidemiology and Environmental Health Joint Research Unit, FISABIO-Universitat Jaume I-Universitat de València, Valencia, Spain. [160]Department of Genes and Environment, Norwegian Institute of Public Health, Oslo, Norway. [161]Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands. [162]Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Queensland, Australia. [163]Generation R Study Group, Erasmus Medical Center, Rotterdam, The Netherlands. [164]School of Women's and Infants' Health, The University of Western Australia, Crawley, Western Australia, Australia. [165]School of Social and Community Medicine, University of Bristol, Bristol, UK. [166]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands. [167]Norwegian Institute of Public Health, Oslo, Norway. [168]CIBER Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain. [169]ISGlobal Barcelona Institute for Global Health, Barcelona, Spain. [170]Division of Metabolic Diseases and Nutritional Medicine, Dr. von Hauner Children's Hospital, Ludwig-Maximilians-University of Munich, Munich, Germany. [171]Telethon Kids Institute, University of Western Australia, West Perth, Western Australia, Australia.

Methods

GWAS meta-analysis. Quality control, imputation and primary association analyses were done using the bioinformatics pipeline Ricopili (see URLs), developed by the Psychiatric Genomics Consortium (PGC)[33]. In order to avoid potential study effects, the 11 PGC samples and the 23 genotyping batches within iPSYCH were each processed separately unless otherwise stated (Supplementary Note).

Stringent quality control was applied to each cohort following standard procedures for GWAS, including filters for call rate, Hardy–Weinberg equilibrium and heterozygosity rates (Supplementary Note). Each cohort was then phased and imputed using the 1000 Genomes Project phase 3 (1KGP3)[34,113] imputation reference panel using SHAPEIT[114] and IMPUTE2 (ref. [115]), respectively. For trio cohorts, pseudocontrols were defined from phased haplotypes prior to imputation.

Cryptic relatedness and population structure were evaluated using a set of high-quality markers pruned for linkage disequilibrium (LD). Genetic relatedness was estimated using PLINK v1.9 (refs [116,117]) to identify first and second-degree relatives ($\hat{\pi} > 0.2$) and one individual was excluded from each related pair. Genetic outliers were identified for exclusion based on principal component analyses using EIGENSOFT[35,118]. This was done separately for each of the PGC cohorts and on a merged set of genotypes for the iPSYCH cohort (Supplementary Note). Across studies, a total of 20,183 cases and 35,191 controls remained for analysis after quality control.

Genome-wide association analyses for the 11 PGC samples and the 23 waves in iPSYCH were performed using a logistic regression model with the imputed marker dosages in PLINK v1.9 (refs [116,117]). Principal components were included as covariates to control for population stratification[35,118], along with relevant study-specific covariates where applicable (Supplementary Note, Supplementary Table 1). Subsequently the results were meta-analyzed using an inverse-variance weighted fixed effects model, implemented in METAL (version 2011-03-25)[36]. Variants were filtered and included if imputation quality (INFO score) was >0.8 and minor allele frequency (MAF) > 0.01. Only markers supported by an effective sample size $N_{eff} = 4/(1/N_{cases} + 1/N_{controls})$[119] >70% were included. After filtering, the meta-analysis included results for 8,047,421 markers.

Conditional analysis. Twelve independent genome-wide significant loci were identified by LD clumping and merging loci within 400 kb (Supplementary Note). In two of these loci, a second index variant persisted after LD clumping. The two putative secondary signals were evaluated by considering analysis conditional on the lead index variant in each locus. In each cohort, logistic regression was performed with the imputed genotype dosage for the lead index variant included as a covariate. All covariates from the primary GWAS (for example, principal components) were also included. The conditional association results were then combined in an inverse-variance weighted meta-analysis.

Genetic correlations between ADHD samples. Genetic correlation between the European ancestry PGC and iPSYCH GWAS results was calculated using LD score regression[37]. The regression was performed using pre-computed LD scores for HapMap3 SNPs calculated based on 378 individuals of European ancestry from the 1000 Genomes Project (see URLs). Only results for markers with an imputation INFO score > 0.90 were included in the analysis. Additionally, a bivariate GREML analysis was conducted using GCTA[120] to estimate the genetic correlation between PGC case/control and trio study designs.

Polygenic risk scores for ADHD. The iPSYCH sample were split into five groups, and, subsequently, five leave-one-out association analyses were conducted, using four out of five groups and the PGC samples as training datasets[38]. PRS were estimated for each target sample using variants passing a range of association *P*-value thresholds in the training samples. PRS were calculated by multiplying the natural log of the odds ratio of each variant by the allele dosage (imputation probability), and whole-genome polygenic risk scores were obtained by summing values over variants for each individual.

For each of the five groups of target samples, PRS were normalized, and the significance of the case–control score difference was tested by standard logistic regression, including principal components. For each target group and for each *P*-value threshold, the proportion of variance explained (Nagelkerke's R^2) was estimated by comparing the regression with PRS to a reduced model with covariates only. The OR for ADHD within each PRS decile group was estimated based on the normalized score across groups (using the *P*-value threshold with the highest Nagelkerke's R^2 within each target group) (Fig. 3). OR was also estimated using logistic regression on the continuous scores for each target group separately, and an OR based on all samples using the normalized PRS score across all groups (Supplementary Fig. 9). Additionally PRS were evaluated in the PGC samples using the iPSYCH sample as training sample, following the approach described above (Supplementary Note).

SNP heritability and intercept evaluation. LD score regression[37] was used to evaluate the relative contribution of polygenic effects and confounding factors, such as cryptic relatedness and population stratification, to deviation from the null in the genome-wide distribution of GWAS χ^2 statistics. Analysis was performed using pre-computed LD scores from European-ancestry samples in the 1000 Genomes Project (see URLs) and summary statistics for the European-ancestry ADHD GWAS to ensure matching of population LD structure. The influence of confounding factors was tested by comparing the estimated intercept of the LD score regression to one, its expected value under the null hypothesis of no confounding from for example population stratification. The ratio between this deviation and the deviation of the mean χ^2 from one (that is it's expected value under the null hypothesis of no association) was used to estimate the proportion of inflation in χ^2 attributable to confounding as opposed to true polygenic effects (ratio = (intercept-1)/(mean χ^2−1)). SNP heritability was estimated based on the slope of the LD score regression, with heritability on the liability scale calculated assuming a 5% population prevalence of ADHD[39].

Partitioning of the heritability. SNP heritability was partitioned by functional category and tissue association using LD score regression[40]. Partitioning was performed for 53 overlapping functional categories, as well as 220 cell-type-specific annotations grouped into 10 cell-type groups, as described in Finucane et al.[40]. For both sets of annotations, we used previously computed LD scores and allele frequencies from European ancestry samples in the 1000 Genomes Project (see URLs).

Additionally, we expanded the cell-type specific heritability analysis by including an annotation based on information about H3K4Me1 imputed gapped peaks excluding the broad MHC-region (chr6:25–35MB), generated by the Roadmap Epigenomics Mapping Consortium[121,122] (Supplementary Note). The analyses were restricted to the European GWAS meta-analysis results to ensure matching of population LD structure. Results for each functional category were evaluated based on marginal enrichment, defined as the proportion of SNP heritability explained by SNPs in the annotation divided by the proportion of genome-wide SNPs in the annotation[40]. For each cell-type group and each H3K4Me1 cell-type annotations, the contribution to SNP heritability was tested conditional on the baseline model containing the 53 functional categories.

Genetic correlations of ADHD with other traits. The genetic correlations of ADHD with other phenotypes were evaluated using LD score regression[42]. For a given pair of traits, LD score regression estimates the expected population correlation between the best possible linear SNP-based predictor for each trait, restricting to common SNPs. Such correlation of genetic risk may reflect a combination of colocalization, pleiotropy, shared biological mechanisms, and causal relationships between traits. Correlations were tested for 211 phenotypes with publically available GWAS summary statistics using LD Hub[41] (Supplementary Note; URLs). Additionally, we analyzed on our local computer cluster the genetic correlation of ADHD with eight phenotypes: human intelligence[103], four phenotypes related to education and cognition analyzed in samples from the UK Biobank[49] (college/university degree, verbal–numerical reasoning, memory and reaction time), insomnia[60], anorexia nervosa[44], and major depressive disorder[43]. The genetic correlation with major depressive disorder was tested using GWAS results from an updated analysis of 130,664 cases with major depressive disorder and 330,470 controls from the Psychiatric Genomics Consortium. As in the previous LD score regression analyses, this estimation was based on summary statistics from the European GWAS meta-analysis, and significant correlations reported are for traits analyzed using individuals with European ancestry.

Credible set analysis. We defined a credible set of variants in each locus using the method described by Maller et al.[123]. (Supplementary Note), implemented by a freely available R script (URLs). Under the assumption that (a) there is one causal variant in each locus, and (b) the causal variant is observed in the genotype data, the credible set can be considered to have a 99% probability of containing the causal variant. For each the 12 genome-wide significant loci, variants within 1 MB and in LD with correlation $r^2 > 0.4$ to the index variant were considered for inclusion in the credible set analysis. The credible set analysis was done using the European GWAS meta-analysis to ensure consistent LD structure in the analyzed cohorts.

Biological annotation of variants in credible set. The variants in the credible set for each locus were annotated based on external reference data in order to evaluate potential functional consequences. In particular, we identify: (a) gene and regulatory consequences annotated by Variant Effect Predictor (VEP) using Ensembl with genome build GRCh37[124]. We exclude upstream and downstream consequences, and consequences for transcripts that lack a HGNC gene symbol (for example vega genes). (b) Variants within 2 kb upstream of the transcription start site (TSS) of at least one gene isoform based on Gencode v19[125]. (c) Variants annotated as interacting with a given gene in Hi-C data from samples of developing human cerebral cortex during neurogenesis and migration[126]. Annotations are considered for both the germinal zone (GZ), primarily consisting of actively dividing neural progenitors, and the cortical and subcortical plate (CP), primarily consisting of post-mitotic neurons. (d) Variants identified as expression quantitative trait loci (eQTLs) based on gene expression in the Genotype-Tissue Expression (GTEx)[127] project database or BIOS[79]. Expression quantitative trait loci were annotated using FUMA (see URLs). We restricted to eQTL associations with

false discovery fate (FDR) $< 1 \times 10^{-3}$ within each dataset. (e) Chromatin states of each variant based on the 15-state chromHMM analysis of epigenomics data from Roadmap[121]. The 15 states summarize to annotations of active chromatin marks (that is Active TSS, Flanking Active TSS, Flanking Transcription, Strong Transcription, Weak Transcription, Genic Enhancer, Enhancer, or Zinc Finger [ZNF] gene), repressed chromatin marks (Heterochromatin, Bivalent TSS, Flanking Bivalent TSS, Bivalent Enhancer, Repressed Polycomb, or Weak Repressed Polycomb), or quiescent. The most common chromatin state across 127 tissue/cell types was annotated using FUMA (see URLs). We also evaluated the annotated chromatin state from fetal brain.

Gene-set analyses. Gene-based association with ADHD was estimated with MAGMA 1.05[88] using the summary statistics from the European GWAS meta-analysis ($N_{cases} = 19{,}099$, $N_{controls} = 34{,}194$; Supplementary Note, Supplementary Table 1). Association was tested using the SNP-wise mean model, in which the sum of -log(SNP *P*-value) for SNPs located within the transcribed region (defined using NCBI 37.3 gene definitions) was used as the test statistic. MAGMA accounts for gene-size, number of SNPs in a gene and LD between markers when estimating gene-based P-values. LD correction was based on estimates from the 1000 Genomes Project Phase 3 European ancestry samples[34].

The generated gene-based *P* values were used to analyze sets of genes in order to test for enrichment of association signals in genes belonging to specific biological pathways or processes. In the analysis only genes on autosomes and genes located outside the broad MHC region (hg19:chr6:25–35 M) were included. We used the gene names and locations and the European genotype reference panel provided with MAGMA. For gene sets we used sets with 10–1,000 genes from the Gene Ontology sets[86] curated from MsigDB 6.0 (ref. [87]).

Targeted *FOXP2* downstream target gene sets were analyzed for association with ADHD. Three sets were examined: (1) putative target genes of *Foxp2* that were enriched in wild type compared to control *Foxp2* knockout mouse brains in ChIP-chip experiments (219 genes), (2) genes showing differential expression in wild type compared with *Foxp2* knockout mouse brains (243 genes), and (3) *FOXP2* target genes that were enriched in either or both basal ganglia (BG) and inferior frontal cortex (IFC) from human fetal brain samples in ChIP-chip experiments (258 genes). Curated short lists of high-confidence genes were obtained from Vernes et al.[89]. and Spiteri et al.[90].

A set of evolutionarily highly constrained genes were also analyzed. The set of highly constrained genes was defined using a posterior probability of being loss-of-function intolerant (pLI) based on the observed and expected counts of protein-truncating variants within each gene in a large study of over 60,000 exomes from the Exome Aggregation Consortium (ExAC)[91]. Genes with pLI ≥ 0.9 were selected as the set of highly constrained genes (2,932 genes).

Replication of GWAS loci. To replicate the results of the ADHD GWAS meta-analysis, we compared the results to those of analyses of cohorts from deCODE and 23andMe, and a meta-analysis of two independent studies conducted by EAGLE and QIMR (referred to as EAGLE/QIMR). We evaluated evidence for replication based on: (a) sign tests of concordance between the ADHD GWAS meta-analysis and each replication cohort; (b) comparison of bias-corrected effect sizes between the ADHD GWAS and the deCODE and 23andMe replication cohorts; (c) genetic correlation between the ADHD GWAS and the 23andMe and EAGLE/QIMR replication cohorts; (d) meta-analysis of the ADHD GWAS meta-analysis results with the results from each replication cohort; and (e) tests of heterogeneity between the ADHD GWAS and each replication cohort.

For the sign test, we first identified the overlapping SNPs present in the ADHD GWAS and each of the three replication analyses (that is deCODE, 23andMe, and EAGLE/QIMR). For each replication cohort intersecting SNPs were then clumped for LD ($r^2 > 0.05$ within 1 Mb) for all variants with $P < 1 \times 10^{-4}$ in the ADHD GWAS (or $P < 1 \times 10^{-5}$ for the deCODE replication) using 1000 Genomes Phase 3 data on European ancestry populations. After clumping, sign tests were performed to test the proportion of loci with a concordant direction of effect in the replication cohort (π) using a one sample test of the proportion with Yates' continuity correction[128] against a null hypothesis of $\pi = 0.50$ (i.e., the signs are concordant between the two analyses by chance) in R[129]. This test was evaluated separately for concordance in deCODE, 23andMe, and EAGLE/QIMR for loci passing *P*-value thresholds of $P < 5 \times 10^{-8}$ (i.e., genome-wide significant loci), $P < 1 \times 10^{-7}$, $P < 1 \times 10^{-6}$, $P < 1 \times 10^{-5}$, and $P < 1 \times 10^{-4}$ in the ADHD GWAS meta-analysis (Supplementary Note).

In addition to testing concordance for the direction of effect, we also evaluated replication for the magnitude of the effect sizes. Specifically, for each of deCODE and 23andMe we regressed the effect size in the replication cohort (that is the log odds ratio) on the estimated effect size from the ADHD GWAS after adjustment for winner's curse for loci with $P < 1 \times 10^{-6}$. Winner's curse correction is performed by computing posterior mean estimates of marginal SNP effects β_j after fitting a spike-and-slab distribution

$$\beta_j \sim \left\{ \begin{matrix} 0 \\ N(0, \tau^2) \end{matrix} \right\} \begin{matrix} \text{with probability } \pi \\ \text{otherwise} \end{matrix}$$

by maximum likelihood as described by Okbay et al.[78]. (Supplementary Note). For the regression of effect sizes we oriented all variants in the direction of the risk increasing allele estimated from the ADHD GWAS, constrained the intercept to zero, and weighted the variants proportional to the inverse of their squared standard error from the ADHD GWAS. A regression slope of one indicates "ideal" replication of all loci in the regression, whereas a slope of zero indicates no replication.

Genetic correlation of the ADHD GWAS with the 23andMe and EAGLE/QIMR results was computed using LD score regression[37] with pre-computed European ancestry LD scores following the same procedure as described above for other genetic correlation analyses. Genetic correlation could not be computed for deCODE since results were only available for top loci from the ADHD GWAS. To further explore the moderate genetic correlation between the 23andMe results and the ADHD GWAS we also evaluated the genetic correlation between traits from 23andMe and traits from LD Hub (URLs)[42]. To evaluate the magnitude of the observed differences in r_g we consider both the absolute difference (that is $|r_{g,ADHD} - r_{g,23andMe}|$) and the test of an approximate *z* score for this difference (Supplementary Note):

$$Z = \frac{r_{g,ADHD} - r_{g,23andMe}}{\sqrt{SE^2_{ADHD} + SE^2_{23andMe}}}$$

We do not expect this to be an ideal formal test for the difference between two genetic correlations, and therefore emphasize caution in interpreting the precise results. Nevertheless, it does offer a useful benchmark for evaluating the magnitude of the difference between the r_g estimates in the context of the uncertainty in those values.

Finally, we meta-analyzed the ADHD GWAS with the results from each replication cohort. For deCODE and 23andMe inverse variance-weighted meta-analyses were performed. For meta-analysis with the EAGLE/QIMR GWAS of ADHD-related behaviors in childhood population samples we used a modified sample size-based weighting method. Modified sample size-based weights were derived to account for the respective heritabilities, genetic correlation, and measurement scale of the GWASs (Supplementary Note). To summarize, given *z* scores Z_{1j} and Z_{2j} resulting from GWAS of SNP *j* in a dichotomous phenotype (for example ADHD) with sample size N_1 and a continuous phenotype (for example ADHD-related traits) with sample size N_2, respectively, we calculate

$$Z_{j,meta} = \frac{\sqrt{\tilde{N}_{1j}} Z_{1j} + \sqrt{\tilde{N}_{2j}} \tilde{Z}_{2j}}{\sqrt{\tilde{N}_{1j} + \tilde{N}_{2j}}}$$

where

$$\tilde{Z}_{2j} = sign(r_g) \frac{Z_{2j}}{\sqrt{1 + (1 - r_g^2) N_{2j} h_2^2 l_j / M}}$$

$$\tilde{N}_{1j} = N_{1j} \frac{P(1-P)\phi(\Phi^{-1}[K])^2}{[K(1-K)]^2}$$

$$\tilde{N}_{2j} = N_{2j} \frac{r_g^2 h_2^2 / h_1^2}{1 + (1 - r_g^2) N_{2j} h_2^2 l_j / M}$$

The adjusted sample sizes $\tilde{N}_1$ and $\tilde{N}_2$ reflect differences in power between the studies due to measurement scale and relative heritability that is not captured by sample size. The calculation of $\tilde{Z}_2$ reduces the contribution of the continuous phenotype's GWAS to the meta-analysis based on imperfect genetic correlation with the dichotomous phenotype of interest (that is ADHD). The adjustments are computed based on the sample prevalence (*P*) and population prevalence (*K*) of the dichotomous phenotype, the estimated liability scale SNP heritability of the two phenotypes (h_1^2 and h_2^2), and the genetic correlation (r_g) between the two phenotypes, as well as the average SNP LD score (l_j) and the number of SNPs (*M*). Heritability and genetic correlation values to compute these weights are computed using LD score regression. This meta-analysis weighting scheme is consistent with weights alternatively derived based on modelling the joint distribution of marginal GWAS beta across traits[130].

To test heterogeneity with each replication cohort, we considered Cochran's *Q* test of heterogeneity in the meta-analyses. Specifically, we evaluated the one degree of freedom test for heterogeneity between the ADHD GWAS meta-analysis and the replication cohort.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The PGC's policy is to make genome-wide summary results public. Summary statistics with the results from the ADHD GWAs meta-analysis of iPSYCH and the PGC samples are available on the PGC and iPSYCH websites (https://www.med.unc.edu/pgc/results-and-downloads and http://ipsych.au.dk/downloads/).

GWA summary statistics with results from the GWAS of ADHD symptom scores analyzed in the EAGLE sample can be accessed at the PGC website (link above). Summary statistics for the 23andMe dataset can be obtained by qualified researchers under an agreement with 23andMe that protects the privacy of the 23andMe participants. For access to genotypes from the PGC cohorts and the iPSYCH sample, interested researchers should contact the lead PIs (iPSYCH, A.D.B.; P.G.C., B.M.N. and S.V.F.).

References

113. Sudmant, P. H. et al. An integrated map of structural variation in 2,504 human genomes. *Nature* **526**, 75–81 (2015).
114. Delaneau, O., Marchini, J. & Zagury, J. F. A linear complexity phasing method for thousands of genomes. *Nat. Methods* **9**, 179–181 (2011).
115. Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes. *G3* **1**, 457–470 (2011).
116. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
117. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
118. Galinsky, K. J. et al. Fast principal-component analysis reveals convergent evolution of ADH1B in Europe and East Asia. *Am. J. Hum. Genet.* **98**, 456–472 (2016).
119. Winkler, T. W. et al. Quality control and conduct of genome-wide association meta-analyses. *Nat. Protocols* **9**, 1192–1212 (2014).
120. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
121. Roadmap Epigenomics Consortium et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
122. Ernst, J. & Kellis, M. Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. *Nat. Biotechnol.* **33**, 364–376 (2015).
123. Wellcome Trust Case Control Consortium. Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat. Genet.* **44**, 1294–1301 (2012).
124. McLaren, W. et al. The Ensembl variant effect predictor. *Genome. Biol.* **17**, 122 (2016).
125. Harrow, J. et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* **22**, 1760–1774 (2012).
126. Won, H. et al. Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* **538**, 523–527 (2016).
127. The GTEx Consortium. The genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
128. Yates, F. Contingency tables involving small numbers and the χ^2 test. *Supp. J. Royal Stat. Society* **1**, 217–235 (1934).
129. R Core Team. R: A language and environment for statistical computing. http://www.r-project.org/ (2014).
130. Turley, P. et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat. Genet.* **50**, 229–237 (2018).

nature research

Corresponding author(s): NG-A46340R

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	Collection of genotypic information was done using SNP array genotyping and subsequent calling of genotypes were done using the softwares GenCall, Birdseed and GenomeStudio
Data analysis	Quality control and association analyses were done using the Ricopili pipeline: https://github.com/Nealelab/ricopili, which include the follwoing software: ShapeIt, IMPUTE2, Plink 1.9, Eigensoft 6.1.3, METAL 2011-03-25. For gene-based and gene-set analyses we used MAGMA 1.06 Analyses of credible SNPs were done using https://github.com/hailianghuang/FMsummary Functional annotation of credible SNPs was done using Ensemble Variant Effect Predictor (VEP). SNPs associated with gene expression were annotated using FUMA (http://fuma.ctglab.nl/) SNP heritability, partitioning of the heritability and genetic correlations were estimated using LD score regression (https://github.com/bulik/ldsc) and LD hub (http://ldsc.broadinstitute.org/). Genetic correlation between PGC case-control and trio samples was calculated using GCTA v1.26.0 Meta analyses of continuous and dichotomous ADHD measures were calculated

using adjusted sample size weighted meta-anaysis, described in details in the Supplementary Information.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Availability of genotype data and summary statistics
For access to genotypes from the PGC cohorts and the iPSYCH sample interested researchers should contact the lead PIs (iPSYCH: lead PI Anders D. Børglum; PGC: Benjamin Neale and Stephen Faraone). Summary statitistics can be downloaded from:
https://www.med.unc.edu/pgc/results-and-downloads
http://ipsych.au.dk/downloads/
http://www.wikigenes.org/e/art/e/348.html

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calcualtion was made. Previous studies of polygenic psychiatric disorders (e.g. schizophrenia) have demonstrated that high numbers of cases and controls (in line with the sample size analysed in this study) yield enough power to detect common risk variants with low effect sizes.
Data exclusions	Within each analysed cohort we aimed at analysing genetically homognous samples of unrelated individuals. Related individuals were excluded based on Identity by State analyses (pseudo controls were used for trios) and genetic outliers were excluded based on principal component analyses
Replication	The results from our primary GWAS of diagnosed ADHD were replicated using results from analyses of three studies: a GWAS meta-analysis of ADHD sympotom scores in the general population (EAGLE and QIMR cohorts), GWAS of selfreported ADHD (the 23andMe sample) and a sample of diagnosed ADHD (deCODE). We evaluated our results in the three replication studies in three ways: 1) sign test 2) genetic correlation 3) meta-analysis of the combined samples. The results of these three replication apporaches support the results from our primary GWAS meta-analysis.
Randomization	Allocation into groups was not random. Individuals were allocated into the case group based on having a diagnosis of ADHD. The controls in each cohort did not have a diagnosis of ADHD.
Blinding	No blinding was done in this study. The design was a case-control study and therefore it was fundamental for the analyses that researchers knew the case-control status of the included individuals

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☐	☒ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Unique biological materials

Policy information about availability of materials

Obtaining unique materials: No unique material were used

Human research participants

Policy information about studies involving human research participants

Population characteristics: *Describe the covariate-relevant population characteristics of the human research participants (e.g. age, gender, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*

Recruitment:

In the meta-analysis we included 11 cohorts form the Psychiatric Genetics Consortium (PGC) and the iPSYCH cohort. In the seperate GWASs we corrected for population stratification using relevant principal components from principal component analyses as covariates.
The iPSYCH cohort was processed in 23 genotyping waves (genotyping, qc and imputation were done seperately for theses batches) of approximately 3,500 individuals each. In order to control for potential batch effects we included “wave” as a covariate in the regression models of all downstream analyses when relevant.
Sex was not used as covariate, as we found no indication of sex being a confounder in our analyses (results not shown).